

# No more protection

Cutting NASA's science budgets is one thing; rejecting the agency's historic role in the study of Earth is something else entirely.

Ten years ago this month, NASA scientists found possible evidence for life on Mars in a meteorite, kick-starting the nascent discipline of astrobiology (*Science* 273, 924–930; 1996). That particular evidence has not stood the test of time, and the infant discipline has also disappointed some.

But the astrobiological vision of a Universe that had a role for biology as well as for chemistry and physics was a powerful one. It linked the study of some 4 billion years of life on Earth with a yearning for hints of life beyond. It was a vision that balanced the outward urge to explore space with an inward appreciation of the sort of world from which we set out and to which we come home. This is the sort of balance that NASA needs today — and which, to judge from a contentious change in the space agency's mission statement, its leadership seems happy to lose.

The agency's current priority, set by President George W. Bush and so far acquiesced to by Congress, is to inaugurate a new age of human exploration. This effort, by no means assured of success, will require cuts in spending elsewhere. Science — never NASA's core concern, nor ever meant to be — has suffered as a result. The most recent potential victim is scientific research on the International Space Station, which the agency seems to have considered putting on hold for a couple of years.

Congress will probably safeguard this particular programme (see page 492), although that will be no great victory. The value of the science slated for the station is not high, and pretty much every scientist outside the microgravity community could imagine better ways of spending the money (not that they will see any of it). It could be cogently argued that, if such cuts help to get the space station finished and the space shuttle grounded for good, they would be a price worth paying. Other shifts at the agency — most noticeably its decision to remove Earth science from the front of its mission statement — are far more disturbing.

## Look back in wonder

President Bush's call for human exploration of the Solar System's more accessible deserts is described as a "vision for space exploration", but this puts the cart before the horse. Although NASA needs direction and leadership, it is a producer of visions as much as their consumer. The Apollo programme, for example, provided a vision of America at its 'can do' best, with an impressive blend of corporate effectiveness, *esprit de corps* and extraordinary technical achievement.

Apollo also played a key role in perhaps the greatest of the visions NASA has given the world — that of Earth from space. No images from any of the Apollo missions speak more powerfully than those of Earth rising above the barren Moon, or of Earth alone in space, the Sahara golden, the Antarctic diamond white, the clouds and seas and surfaces between them ineffably rich and complex. Winning a

new point of view beyond the surface of Earth has provided nothing aesthetically, spiritually or intellectually more important than the ability to look back at our planet below. NASA, as its more poetically inclined supporters have pointed out before, thus embodies the wisdom of T. S. Eliot's *Little Gidding*:

*We shall not cease from exploration  
And the end of all our exploring  
Will be to arrive where we started  
And know the place for the first time.*

This vision from outside is not only inspiring; it is also empowering, providing a fresh way of studying the Earth. As the population dependent on Earth's resources grows, and its perturbations of the planet's atmosphere and ecosystems escalate, studies of the processes being perturbed become ever more valuable.

## Mission impoverished

That is why recent changes to NASA's mission statement are more disturbing than any single cut could be. The goals in the mission statement, developed through a thorough set of consultations earlier this decade, used to begin "to understand and protect our home planet". This imperative is now gone, replaced by a commitment to "pioneering the future".

NASA administrator Michael Griffin argues that the agency is still committed to Earth science through the part of its mission statement that commits it to "scientific discovery", and that what "protect" meant was never clear. For all that, the excision still echoes the only decent scene in the movie *Superman III*: "I hope you don't expect me to save you," that other icon of the American way informs a damsel in distress, "cause I don't do that any more." Except this time there's no artificial kryptonite to pin the blame on. And it's not funny.

It is bad enough that Earth science at NASA has already fallen victim to cuts and cancellations — as has, for what its worth, astrobiology. Now an important rhetorical basis for resisting more attrition has been removed, feeding fears that a real understanding of how the climate works is not high on the administration's agenda. Earth sciences are still well represented in NASA's plans, but they have been symbolically set aside to further a vision that looks only outwards, never back.

This situation should not be allowed to continue. Employees, grantees and others with a stake in the agency should press hard to have the change reversed. A NASA that does not see its interest in the living Earth as essential is as much of a betrayal as a Superman without altruism. ■

**"I hope you don't expect me to save you, 'cause I don't do that any more."**



# A foundation for Africa

A research plan that has to be seen to be believed.

**N**igeria's plan to set up a multi-billion-dollar endowment for science (see *Nature* **442**, 334; 2006) will be a tremendous boon for the development of science in Africa's most populous nation — if it comes to fruition.

President Olusegun Obasanjo's vow to set up an agency based on the US National Science Foundation (NSF) to distribute the endowment's surplus is welcome and far-sighted. Such an agency is necessary to ensure that the money is sensibly spent. The structure of the proposed agency will tell us whether its founders are serious about replicating the independence that is the NSF's hallmark.

The idea may not be as far-fetched as it sounds. The NSF was itself ingeniously designed to protect scientific funding decisions from the propensity of politicians to treat research cash as a slush fund to be shared out among the most powerful — exactly the problem that any attempt to support science in Nigeria will face.

The characteristics of the US agency are easy to recognize. It is managed by a board of prominent scientists, who serve for terms of six years. One-third of them rotate off every two years, so the board cannot readily be 'stacked' by an incoming US president. The director serves for up to six years too, unlocking the appointment from the president's four-year election cycle.

The staff of the agency is small and professional, and includes academics on secondment from universities. The staff set up advisory

panels, often including international representation, in different subdisciplines and make decisions on grants. Nearly all the agency's budget goes on grants: there are no laboratories or permanent infrastructure to swallow the cash.

There are issues with transplanting this widely admired model, however. It works best where university professors are already on good salaries, as in the United States, and only need research grants to cover the direct costs of research itself.

UNESCO, which is working with Nigeria on its science policy, has suggested that about five of Nigeria's existing universities be selected for extra funding. That proposal, along with the endowment itself, will be considered by Nigeria's parliament this autumn. The president's advisers say they would like Nigeria to provide the bulk of the investment, with some assistance from foreign donors.

It may sound odd for Nigeria, with its buoyant oil revenues, to expect hard-pressed donor agencies to fund such a plan. But organizations such as the World Bank often invest in fostering long-term economic development — and adding to the fund would give donors some say in how it is run.

Setting up a politically independent science agency with an endowment to ensure its long-term viability would be an eminently sensible use of windfall oil revenues. Long-term observers of the nation's politics will no doubt be sceptical. But if Obasanjo can pull it off before his expected departure next year, he would leave a spectacular legacy for science in Nigeria, and in all of Africa. ■

**"President Obasanjo's vow to set up an agency based on the NSF to distribute the endowment's surplus is welcome and far-sighted."**

# What's in a name?

Chemistry is here to stay.

**I**n this week's issue, prominent chemists identify six important questions currently being asked on the frontiers of their field (see page 500). The calibre of the selection demonstrates clearly that the discipline remains alive with fresh ideas and challenges.

However, chemistry's poor public image, dogged by echoes of industrial pollution and the pejorative connotations of the word 'synthetic', has led to suggestions that it rebrand itself with a sexier label — molecular sciences, perhaps. The idea is not just public-relations flannel: it is also a response to the undoubted movement of the discipline's centre of gravity since Johann Hartmann took the first university chair of chemistry at Marburg in Germany in 1609.

There are plenty of precedents. In recent decades, many geologists have become 'Earth scientists', some metallurgists are now materials scientists, and biology departments have splintered into all manner of subdivisions. Most of these renamings are down to more than fashion; they reflect a genuine shift in emphasis.

But there are strong arguments that 'chemistry' remains the best word to describe the sciences of matter and its transformations. Far from being an endangered subject, chemistry is a victim of its own success. It has given us the tools and concepts, for example, to probe questions in the life sciences that were previously impenetrable.

As a label, 'biochemistry' fails to do justice to this encroachment of chemistry into biology, being associated primarily with the study of enzyme kinetics. Similarly, chemists can now design materials from the atomic level upwards, whether they work in departments of chemical engineering, polymer science or materials. Chemists may even give electronic engineers a run for their money in creating self-assembling circuits and memories.

The problem is that so little of the credit falls on chemistry itself. 'Chemical biology', for example, presents the chemical aspect as an adjunct to another discipline, while a great deal of inventive chemistry comes under the umbrella of nanotechnology. This tendency is already reflected in university reshuffles, at Harvard and elsewhere, that might displace parts of chemistry to other departments.

Such rebranding is not the way to secure chemistry's position as an independent discipline. Neither will that be achieved merely by loudly trumpeting chemistry's considerable achievements. Chemistry needs, instead, to reassert itself as a core scientific discipline — albeit one that is inherently interdisciplinary and has a strong applied component. Perhaps it is time for chemistry departments to rethink the subject's internal structure: the traditional divisions of physical, organic and inorganic chemistry have long since become irrelevant in many respects.

Another challenge for chemists is to be ambitious in defining the discipline's boundaries, particularly in relation to the life sciences. And in terms of public perception, chemistry has to find a way of establishing its centrality in the mysterious process we call life. ■

# RESEARCH HIGHLIGHTS

## Oily snack

*Nature Biotechnol.* doi:10.1038/nbt1232 (2006)

A marine bacterium that guzzles oil has had its genome sequenced, offering the first complete guide to how such bacteria thrive. This should help environmental scientists work out how to promote the growth of such bacteria in oil spills, so speeding the pollution's clean up.

Vitor Martins dos Santos of the German Research Centre for Biotechnology in Braunschweig and his colleagues studied *Alcanivorax borkumensis*, a microbe that feeds exclusively on oil. Its genome contains sequences that encode surfactants, for example. Part water- and part oil-soluble, these surfactants would help break up the oil. The researchers also identified previously unknown enzymes that might have applications in industry.



J. MORGAN/ALAMY

## NEUROBIOLOGY

### Enhanced sense of smell

*Cell* **126**, 403–413 (2006)

Some remarkable cross-chromosome canoodling explains how each neuron in the mouse nose manufactures only one of the 1,300 receptor proteins for detecting smell, according to new research.

Using fluorescent labels, Richard Axel and his colleagues at Columbia University, New York, show that a known DNA 'enhancer' region on one chromosome is associated with the single, active copy of a smell receptor gene on a different chromosome.

The researchers propose that the control region randomly interacts with and switches on just one receptor gene in each neuron. Mice genetically engineered to carry extra enhancers manufactured more receptors.

This type of gene regulation might also control the expression of other gene families.

## CIRCADIAN BIOLOGY

### Clock of ages

*Genes Dev.* **20**, 1868–1873 (2006)

The body's internal clock is linked to how quickly we age, according to researchers in the United States.

Marina Antoch and Roman Kondratov of the Cleveland Clinic Foundation, Ohio, and their colleagues studied mice lacking the *Bmal1* gene, which encodes a core component of the circadian clock. The

mice lived, on average, only half as long as normal mice and showed signs of premature ageing, such as muscle wasting, cataracts and decreased hair growth.

The accelerated ageing is likely to be caused, in part, by the fact that the mice were less able to neutralize free radicals, highly reactive molecules that damage tissue.

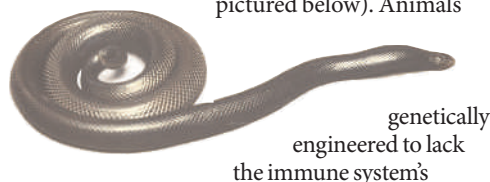
## IMMUNOTOXICOLOGY

### Once bitten...

*Science* **313**, 526–530 (2006)

Our immune system has a hitherto unheralded defence against the potentially deadly effects of toxic venom, biologists have discovered.

Researchers led by Stephen Galli of Stanford University School of Medicine in California injected mice with venom from the deadly burrowing asp (*Atractaspis engaddensis*, pictured below). Animals



mast cells succumbed far more easily to the killer toxin.

Mast cells, the team shows, contain protein-digesting enzymes to combat toxic compounds, forming a crucial line of defence against snake bites and bee stings. The result is a surprise because mast cells, which can also trigger widespread and potentially

damaging inflammation, had previously been viewed as the villain rather than the hero in the body's response to a bite.

## ASTRONOMY

### Explosive measuring tape

*Mon. Not. R. Astron. Soc.* **370**, 773–783 (2006)

Stellar explosions that astronomers had assumed to be clones of each other might stem from two different kinds of star, suggest Filippo Mannucci of the INAF Institute of Radio Astronomy in Florence, Italy, and his colleagues. This could mean that properties of the explosions, such as brightness, differ between them, which would affect how astronomers use these type 1a supernovae to measure distances in space.

The researchers studied how the rate at which supernovae happen depends on the age of the universe and the colour and nature of the galaxy in which they go off. They argue that these three data sets are best fitted by a model with two classes of type 1a supernovae: 'prompt' and 'tardy', relative to when the star formed — an idea they suggest should be further tested.

## MICROSCOPY

### Time to relax

*Proc. Natl Acad. Sci. USA* **103**, 11440–11445 (2006)

Clever physics has helped Stefan Hell and his colleagues to picture tiny cellular features, such as the contents of compartments called endosomes, which are beyond the reach of

A. SHOUB, E. KOCHVA



conventional light microscopes.

The team at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany, had already developed a trick to boost the resolution of fluorescence microscopy from 200 nanometres — the diffraction limit — to around 60 nanometres (*Nature* **440**, 935–939; 2006).

Now the researchers show that pulsing the light used to excite the fluorescent dye can sharpen the resolution further. The pulsing pattern gives fluorescent molecules that would otherwise have become exhausted (a process known as photobleaching) time to relax, allowing the team to ramp up the laser's power and narrow the excitation beam. This made it possible to resolve features only 15–20 nanometres across.

## MATERIALS CHEMISTRY

### Designer glasses

*J. Am. Chem. Soc.* doi:10.1021/ja063353s (2006)

Researchers in Canada have figured out how to frustrate small organic molecules so that they won't form crystals. Instead, the molecules they design should form disordered glasses, useful for their optical properties.

Small organic molecules have been widely explored in crystal engineering, because they can be designed to self-assemble into specific crystalline arrays guided by intermolecular forces such as hydrogen bonding. Such designer crystals might be useful as molecular sieves or catalysts, for example.

James Wuest and his colleagues at the University of Montreal, Quebec, looked at how hydrogen bonding might, in contrast, be used to prevent crystallization. They devised molecules that form aggregates, which cannot easily be closely packed into regular arrays, and so do not crystallize.

## VISION

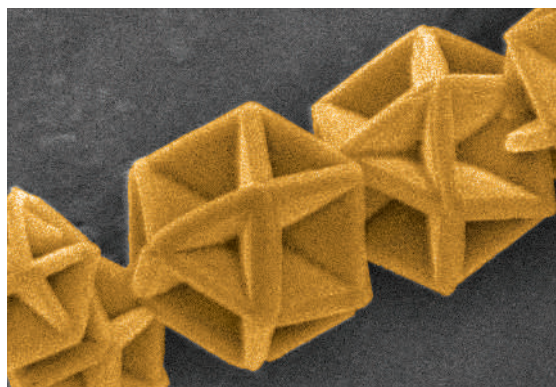
### Are we missing something?

*PloS Biol.* **4**, e254 (2006)

Mammals are missing a photoreceptor protein found in other vertebrates, a new study has shown.

Vertebrate vision relies on opsins — proteins that transform light signals into electrical current in the outer retina. In the inner retina, photoreceptors that contribute to non-visual light responses, such as pupil constriction, depend on an opsin called melanopsin.

Jim Bellingham of the University of Manchester, UK, and his colleagues report that vertebrates other than mammals have two distinct melanopsin genes, rather than



ACS

just one as was previously believed. The question now is: what are the functional consequences of having only one melanopsin? Until we find out, we won't know what we're missing.

## CRYSTAL GROWTH

### Star quality

*Chem. Mater.* doi:10.1021/cm060956u (2006)

Recipe for making geometric 'stars': dissolve copper nitrate in ethylene glycol, add sulphur and bake well. The result, report Shu-Hong Yu of the University of Science and Technology of China in Hefei and his co-workers, is microscopic crystals of copper sulphide that have a beautiful cuboctahedral form (pictured above), reminiscent of the cages drawn by M. C. Escher in his 1948 engraving *Stars*. They are composed of four intersecting hexagonal plates and contain 14 concave cavities.

## RNA INTERFERENCE

### Unwanted effects

*J. Neurosci.* **26**, 7820–7825 (2006)

Using RNA interference in brain cells can severely perturb the cells' function, report researchers from Harvard Medical School in Boston, Massachusetts. This reinforces concerns about side effects of the gene-silencing technique, highlighted in a number of recent studies.

In this work, Bernardo Sabatini and his colleagues added small RNA molecules known as short hairpin RNAs (shRNAs) to rat brain slices. Such shRNAs will silence any gene with a sequence that matches their own — through RNA interference. In this experiment, however, the chosen shRNAs targeted a sequence not found in the rat's genome, so they should have had no effect. But the treatment wreaked havoc, destroying or disturbing the connections between neurons.

The researchers think the damage was caused by an immune response.

## JOURNAL CLUB

Ruth McKernan  
Pfizer, Sandwich, UK

**Pfizer's UK head of Discovery Biology is reminded that drug discovery is not all about science.**

When I read a recent paper by Louis Staudt of the National Cancer Institute in Bethesda, Maryland, and his colleagues, it brought a lump to my throat.

As a drug-company researcher, I am always looking for better ways to find new drug targets. The paper (V. Ngo *et al.* *Nature* **441**, 106–110; 2006) described an elegant RNA-interference approach to identifying critical genes and pathways in B-cell lymphoma, a cancer of the lymph system.

But the work grabbed me for other reasons. It wasn't that *CARD11* — the gene that the team focuses on — offers a new approach to developing drugs for this kind of cancer. Or even that it set me thinking about how to modify the method, for application in other therapeutic areas. As a neuroscientist, I wondered whether I could use a similar approach to find genes that prolong the life of particular neurons in neurodegenerative disorders such as Parkinson's or Alzheimer's disease.

What distinguished this paper from others, for me, was the powerful reminder that drug discovery is not just about science — it's about people too.

When my father died of a B-cell neoplasm six years ago, just before the first draft of the human genome was published, we had little idea of the genes involved, far less any hope that a selective and specific treatment might one day be possible.

Behind every hopeful gene target are families such as mine. I know from experience that most approaches will fail, yet this paper and others like it represent hope for patients, and mark the pace of scientific progress.

So, when I finished reading it, I swallowed hard and returned to my work with a renewed sense of vigour.

## NEWS

# Maths 'Nobel' rumoured for Russian recluse

Some speculation always precedes the announcement of the Fields medals, the most illustrious awards in mathematics. But this year rumours have an extra dimension: one of the prizes to be awarded at the International Congress of Mathematicians in Madrid on 22 August may be given for the solution of a century-old problem.

Favourite to win a medal is Russian mathematician Grigory 'Grisha' Perelman, for his work on the Poincaré conjecture. Three papers now suggest that his proof of the conjecture is right. If it stands up to two years of scrutiny, Perelman will also be eligible for at least a share of a million-dollar mathematics prize (see 'Clay feats').

"I am completely convinced that Perelman has proved the Poincaré conjecture," says John Morgan, a topologist at Columbia University in New York and co-author of the most recent paper analysing Perelman's work.

The Poincaré conjecture is a seemingly simple statement about the nature of three-dimensional surfaces, but it had resisted proof since French mathematician Henri Poincaré put it forward in 1904.

In 2002, Perelman posted to an online preprint server a paper<sup>1</sup> that claimed to "give a sketch of an eclectic proof of this conjecture". By this, Perelman meant not just the Poincaré conjecture, but the geometrization conjecture, a much broader theorem of which Poincaré's statement is a special case.

The geometrization conjecture classifies all the possible forms that three-dimensional surfaces, called three-manifolds, can take. The Poincaré conjecture describes the properties of a subset known as 'simply connected' manifolds — in particular, those for which every loop drawn on the surface can be shrunk to a point.

To tackle the conjecture, Perelman used a mathematical tool known as Ricci flow, developed by Richard Hamilton of Columbia University. Perelman, formerly of the Steklov Institute in Moscow, advanced Hamilton's work by finding a way to deal with certain problems encountered in using Ricci flow to study the deformation of surfaces.

The advent of a possible proof shook a small contingent of mathematicians. "For people studying topology, the whole landscape has changed. It's like waking up one morning after an earthquake," says Bruce Kleiner of Yale University in New Haven, Connecticut.

Perelman followed his first article with two further papers, and in the spring of 2003 he toured the United States to lecture about his work<sup>2</sup>. He has since retreated from the public eye, leaving other mathematicians to comb through his papers line by line, filling in details and searching for holes in his logic.

"How do you decide whether anything is correct?" asks John Ball, president of the International Mathematical Union and chair of the Fields medal committee. "All that can happen

**"For people studying topology, the whole landscape has changed."**



C. CHANG/DAILY PRINCETONIAN

is that clever people, experts in the area, read it and come to an opinion."

Three pairs of respected mathematicians have produced written accounts filling in the details of Perelman's work.

Kleiner and John Lott of the University of Michigan, Ann Arbor, posted their most recent paper to the arXiv preprint server on 25 May<sup>3</sup>. "There was a moment when I really thought there was a problem with the argument," says Kleiner. "But that state of mind

## Bird flu not set for pandemic, says US team

Researchers have tried to create a pandemic H5N1 influenza strain — and failed.

Simply mixing genes from an H5N1 bird-flu strain with those from an H3N2 human strain did not result in a strain that was readily transmissible, at least among ferrets. The scientists who conducted the work, at the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, say it suggests that the

H5N1 virus will require a complex series of genetic changes to evolve into a pandemic strain.

"These data do not mean that H5N1 cannot convert to become transmissible from person to person," says Julie Gerberding, director of the CDC. "We are not out of the woods on pandemic preparedness yet."

Others agree, pointing out that there are many ways a pandemic

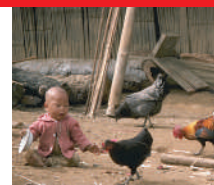
strain could evolve. For instance, strains other than those used in the experiments could get together.

"They need to look at other viruses, because both human and avian flu continue to evolve," says Frederick Hayden, a flu specialist at the University of Virginia, Charlottesville, who is currently working with the World Health Organization.

Since 1997, millions of domestic

and wild birds have died owing to the H5N1 strain of flu. It has infected at least 232 people and killed 134 of them. Scientists are worried that H5N1 will learn to pass easily between humans and kill millions more. In 1957 and 1968, pandemic strains of flu seem to have emerged when bird and human flu viruses exchanged genes, allowing the bird-flu virus to be easily transmitted between people.





### INDONESIAN FLU CASES GO UNSTUDIED

No sequence data have been acquired from birds for nearly a year.

[www.nature.com/news](http://www.nature.com/news)

## Clay feats: the million-dollar questions

In 2000, the privately funded Clay Mathematics Institute in Cambridge, Massachusetts, announced a prize of \$1 million for the solution of each of seven mathematical conundrums — including the Poincaré conjecture.

"I didn't think that I would live to see the Poincaré conjecture solved," says James Carlson, president of the institute. "But we're there, or almost there."

The rules require that the proof appear in a peer-reviewed outlet and then survive two years of scrutiny. Grigory Perelman's papers containing a possible proof of the Poincaré conjecture



Great foundations: the Clay Mathematics Institute.

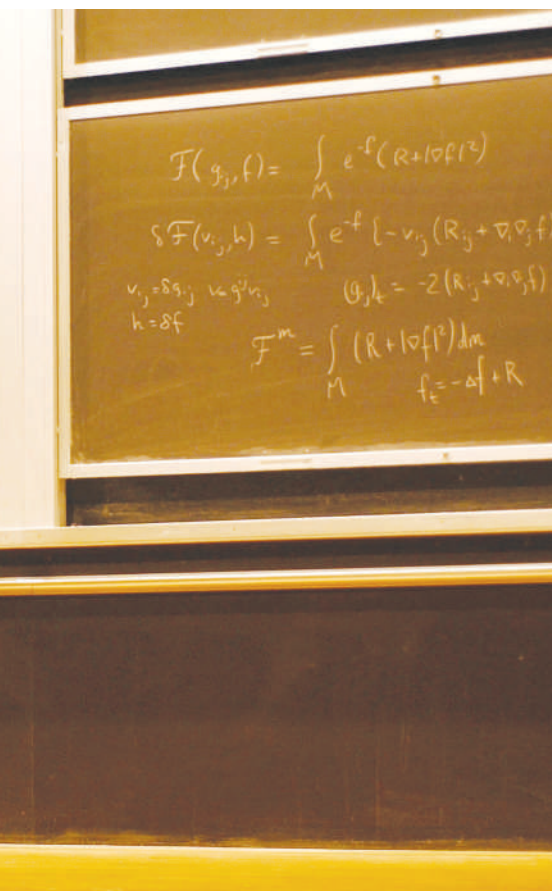
appear online but have not been formally reviewed. Those who have tried to persuade Perelman to publish say that he shows no interest. Other recent papers on his work by different mathematicians may count instead, says Carlson. "These

serve to start the clock on the waiting period," he says.

Some mathematicians suggest that the prize money should be shared between Perelman and Richard Hamilton, a mathematician at Columbia University in New York, who laid much of the groundwork. Those who have followed up the proof could also theoretically lay claim to the money.

Bruce Kleiner of Yale University says he will not seek a share. "I'm not saying that it was easy to fill in the details," he says. "But you can't compare it to the level of ingenuity required to sketch out the proof." **J.H.**

CLAY MATHEMATICS INST.



Board meeting: Grigory Perelman presents his proof of the intractable Poincaré conjecture.

persisted for only a week or two before the issue was resolved."

A 473-page paper<sup>4</sup> by Morgan and Gang Tian of the Massachusetts Institute of Technology appeared on 25 July.

The third paper, from Huai-Dong Cao of Lehigh University in Bethlehem, Pennsylvania, and Xi-Ping Zhu of Zhongshan University

in Guangzhou, China, was published in June<sup>5</sup>. It claims to complete the proof of both conjectures, rather than simply fleshing out Perelman's work. This is justified, argues Shing-Tung Yau, an editor-in-chief on the journal, because Cao and Zhu follow a somewhat different argument to Perelman.

The appearance of these three papers has fuelled rumours that Perelman will receive a Fields prize. Up to four medals are awarded every four years to mathematicians no older than 40, making Perelman just eligible for this year's crop. Not that he will attend the Madrid meeting: the organizing committee received no reply when they invited him to give a plenary lecture.

Even so, many mathematicians think Perelman will be honoured with a medal. "I hope that they give him one. This is such a major

achievement in mathematics, it would be funny if they didn't," says Morgan. There is also a precedent. Two previous Fields medals have been awarded for proofs of an equivalent of the Poincaré conjecture in higher dimensions — to Stephen Smale in 1966 and Michael Freedman in 1986.

And in 2006? Morgan, along with 4,000 guests, will find out in a few weeks — when the medals are announced at the opening ceremony of the congress in Madrid. ■

Jenny Hogan

1. Perelman, G. Preprint at <http://arxiv.org/abs/math.DG/0211159> (2002).
2. Singer, E. *Nature* **427**, 388–389 (2004).
3. Kleiner, B. & Lott, J. Preprint at <http://arxiv.org/abs/math.DG/0605667> (2006).
4. Morgan, J. W. & Tian, G. Preprint at <http://arxiv.org/abs/math.DG/0607607> (2006).
5. Cao, H.-D. & Zhu, X.-P. *Asian J. Math.* **10**, 165–492 (2006).

To test whether H5N1 might do this, the CDC scientists used a technique called reverse genetics to snip genes out of H3N2 and H5N1 viruses and recombine them into hybrid bird-human viruses. They infected ferrets with the hybrids and tested whether the animals got sick and transferred the viruses to other ferrets (T. R. Maines *et al. Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0605134103; 2006).

Ferrets infected with hybrid viruses did not get as sick as those infected with the original H5N1

virus. In addition, the ferrets did not pass the hybrid virus easily to others. And even if they did pass on the virus, the other ferrets did not become fatally ill.

**"They need to look at other viruses, because both human and avian flu continue to evolve."**

The findings seem to indicate that the recombined viruses were less deadly than the original H5N1 strain and unlikely to transfer to

other animals, the scientists say. They hope to repeat the experiments to test the pandemic potential of other viruses — including those taken from patients after 1997, the year that the H5N1 strain they used was isolated.

"We believe this model is a good tool to assess the potential of H5N1 viruses to cause a pandemic in the future," says Jacqueline Katz of the CDC, who led the work.

Many questions remain unanswered, however. The ferret model may not perfectly replicate

human disease, say scientists not involved in the experiments. Nor does the study address whether H5N1 could evolve into a pandemic strain by accumulating mutations if it passed through many people. It also did not test hybrids with human flu viruses other than H3N2.

"The attention being paid to pandemic preparedness is certainly appropriate, and the results shouldn't dissuade people from continuing to progress in that area," says Hayden. ■  
Erika Check

## ON THE RECORD

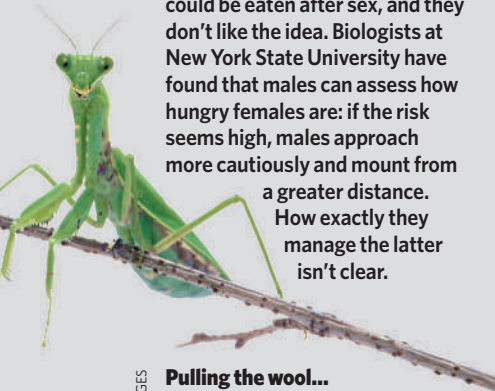
**“We need to turn scientists back into the rock stars they are.”**

Morgan Spurlock is making a movie of *The Republican War on Science*.

## ZOO NEWS

## Hope and prey

In case you were wondering, male praying mantises do know they could be eaten after sex, and they don't like the idea. Biologists at New York State University have found that males can assess how hungry females are: if the risk seems high, males approach more cautiously and mount from a greater distance. How exactly they manage the latter isn't clear.



NOBUI IIDA/GETTY IMAGES

## Pulling the wool...

The trial of disgraced stem-cell scientist Woo Suk Hwang continues to amaze. He is accused of accepting 2 billion won (US\$2.1 million) in private donations, but Hwang insists none of it went to him personally. He told the court last week that part of the money was used in an attempt to clone mammoths.

## NUMBER CRUNCH

How fast does the eye transmit information to the brain?

**6** bits of information, at least, are sent every second by each cell in a guinea pig's optic nerve.

**875,000** bits are calculated to be sent each second by the whole nerve ( $10^5$  cells).

**8,750,000** bits would thus be sent every second by a human optic nerve... about the same as a standard Ethernet connection.

## SCORECARD



## Russian rockets

A Dnepr rocket, based on the Soviet RS-20 missile, has crashed carrying 18 satellites, including 14 'microsats' made by students. NATO dubbed the original missile Satan; it's now being called worse names in classrooms around the world.

Sources: Time, Am. Nat., Curr. Biol.

# NASA threatens to axe science on space station

## WASHINGTON DC

In a move poorly received by key members of the US Congress, NASA last week said that it might shut down scientific research on the International Space Station for at least a year, to save money.

NASA is struggling to fund its mission to send astronauts back to the Moon, while spending around \$200 million a year on research in the station.

But Congress has mandated, as part of the 2005 NASA Authorization Act, that the agency should carry out “to the maximum extent practicable, basic, applied, and commercial research aboard the International Space Station”. The act also specifies that NASA should allocate at least 15% of the funds budgeted for space-station research to experiments not directly related to the human exploration programme.

**“I can't believe they are discussing this with a straight face.”**

“The International Space Station is first and foremost a space laboratory,” says Katie Boyd, a spokeswoman for Senator Richard Shelby (Republican, Alabama). His state hosts the Marshall Space Flight Center, which manages science work aboard the space station. “To suggest that NASA would not take advantage of this unique laboratory is shortsighted.”

A NASA spokesman, Grey Hautaluoma, said that he could not comment on ongoing budgetary discussions. “Right now, that funding is still there,” he says.

About three dozen experiments are currently under way on the station, from biology set-ups that use the astronauts as guinea pigs, to aeronautics experiments that use polyhedrons to test flight-control algorithms.

A year-long hiatus from research would mean more than missing data points. Restarting stalled projects will require staff to be rehired and equipment replaced. In the end, taking time off from research would probably cost more than it saved.

“I can't believe that they would discuss this with a straight face,” says former NASA employee Keith Cowing, who broke the story on his website, NASA Watch.

Some of the 16 other partners in the station agree that the idea doesn't have legs. “If they were really going to, I think they would announce it to us first, before the press,” says Marc Heppener, head of the station science programme at the European Space Agency. ■

Heidi Ledford

With additional reporting by Geoff Brumfiel and Jenny Hogan

NASA



Lost in space: the space station is using up cash that NASA needs to fund its Moon mission.

## The proof is in the product

It's pretty standard to find syntheses in a chemistry journal. But a paper published online last week by the international edition of *Angewandte Chemie* has got the world of organic chemistry humming. It is the second time this year that the journal has published a complete synthesis for a molecule called

hexacyclinol — yet each paper claims a different structure for its final product.

Hexacyclinol is extracted from a mushroom and has some medical potential. The first synthesis of the molecule was published in February by chemist James La Clair of the Xenobe Research Institute in San

Diego (J. J. La Clair *Angew. Chem. Int. Ed.* 45, 2769–2773; 2006).

But La Clair's work raised a few eyebrows. For instance, he said his 37-step synthesis made 3.6 grams of hexacyclinol. “I've never heard of a synthesis of that length producing that much material,” says Larry Overman, an organic chemist at





**TESTOSTERONE DOPING**  
How do the tests work, and are they reliable?  
[www.nature.com/news](http://www.nature.com/news)

J. GUEZ/AFP/GETTY

# Singapore pulls plug on US collaboration

The Singaporean government is known for its generosity in pumping money into international research projects. But it can apparently be ruthless if these projects do not please it. The city-state is shutting down a medical research arm of Johns Hopkins University in Singapore, claiming it has not delivered as planned.

The decision has sent shock waves through other universities and research institutes, some of which are in the initial stages of collaborating with Singapore.

On 20 June, the 60 faculty members and staff of the Division of Biomedical Sciences, Johns Hopkins Medicine in Singapore, were given official notice that as of 1 June 2006, the facility was being wound down. The process will take 12 months, with researchers and staff receiving salaries until the facility closes on 31 May 2007.

"We thought it unwise to continue putting money into something that is not working," says Andre Wan, director of the Biomedical Research Council at the Agency for Science, Technology and Research (A\*STAR). He claims that Johns Hopkins failed to meet key requirements in its contract.

The split is the first major failure of Singapore's recent push into biomedical research, which includes building the ambitious Biopolis research centre.

Singapore and Johns Hopkins Medicine first got together in 1998, to set up Johns Hopkins Singapore as the university's base for medical research, education and clinical studies in southeast Asia. In 2003, the two parties restructured the organization, making it the first full division of Johns Hopkins Medicine outside its US home in Baltimore, Maryland.

Under the five-year contract begun in February 2004, A\*STAR agreed to provide S\$75 million (US\$47.5 million) to cover salaries, facilities



L. ASCUI/REUTERS

Singapore has the funds for grand research ventures such as Biopolis, but it demands results.

and research equipment. For its part, according to A\*STAR, Johns Hopkins aimed to build a centre of immunology, experimental therapeutics and cancer research, to establish a PhD programme and to bring senior researchers with international reputations to Singapore.

Wan alleges that Johns Hopkins has failed to meet 8 out of 13 key performance indicators in the past two years. For example, it agreed to hire at least 12 full-time senior researchers with residence in Singapore by the end of the second year. So far, Wan says it has hired seven, only one of whom fulfils all the requirements.

Gary Stephenson, a spokesman for Johns Hopkins Medicine, declined to comment, saying the university plans to issue a joint statement with A\*STAR soon. But on 22 July, a

Singaporean newspaper reported a Johns Hopkins official as alleging that Singapore failed to meet its financial and educational obligations under the agreement; A\*STAR denies those claims.

The break-up has reminded other universities not to take their collaborative projects with Singapore for granted. For instance, in 2005 Duke University School of Medicine in Durham, North Carolina, established a graduate medical school with the National University of Singapore, as part of a seven-year contract. "The major lesson I'm taking from this is to seek great clarity of how performance in our exciting new venture will be defined," says Sanders Williams, founding dean of the new school.

Ichiko Fuyuno

the University of California, Irvine. "There was quite a bit — a lot — of scepticism in the community."

The sceptics included Scott Rychnovsky, also of Irvine. La Clair claimed that the structure of his compound matched a prediction made in 2002. But Rychnovsky worked out that hexacyclinol should resemble another mushroom compound, panepophenanthrin.

Rychnovsky teamed up with John Porco of Boston University, who

came up with a new synthesis (J. A. Porco Jr *et al.* *Angew. Chem. Int. Ed.* doi:10.1002/anie.200602854; 2006).

The nuclear magnetic resonance (NMR) spectrum for the product matched the 2002 data, suggesting the team had made hexacyclinol. And the crystal structure confirmed their new prediction about its structure.

Which leaves chemists wondering how La Clair got results confirming the original — apparently wrong — structure. And how such a

controversial synthesis made it into a peer-reviewed journal. "I can't see any legitimate explanation for the original work," says Rychnovsky.

La Clair now says that he and Porco may have made two different molecules that happen to have very similar NMR spectra. He has no plans to retract his paper, but has offered to change the name of the compound he says he synthesized.

Because La Clair is unaffiliated with an institution other than the privately

funded Xenobe, there is no obvious body to investigate what happened.

Peter Göllitz, editor of *Angewandte Chemie*, says La Clair's paper was recommended by three well-qualified reviewers. He declined to comment on whether the paper will be retracted. But he says he invited Rychnovsky to publish "so that the discussion and clarification can be done at the place where the original paper was published".  
Emma Marriss

**DDT RETURNS**

The infamous pesticide is on its way back to Africa to fight malaria

[www.nature.com/news](http://www.nature.com/news)

A. JOE/AFP

# Mouse data hint at human pheromones

SELF/LAMY



On that dream date, something really might be in the air. Results from a mouse study may bolster the evidence for human pheromones, the long-debated chemical signals thought to unconsciously sway our behaviour.

A team at the Fred Hutchinson Cancer Research Center in Seattle, Washington, has identified a class of protein receptor in the lining of mouse noses that may also operate in humans.

Researchers know that certain chemicals in mouse urine can alter the social or sexual behaviour of other mice. These chemicals work partly by binding to receptors in a particular structure in the mouse nose, known as the vomeronasal organ. In humans, however, this organ is thought to be defunct and the role of pheromones is hotly debated.

Finding such receptors in the lining of the nose, rather the vomeronasal organ, is a more direct parallel with humans. Stephen Liberles and Linda Buck report their finding online this week in *Nature* (S. D. Liberles & L. B. Buck *Nature* doi:10.1038/nature05066; 2006). They isolated a group of receptors that can be triggered by at least one known mouse pheromone.

Genes encoding this family of receptors are also found in humans. "It's probably our best bet yet for functional pheromone genes in humans," says Timothy Holy, a neurobiologist at Washington University in St Louis, Missouri.

For their part, Liberles and Buck are cautious about labelling the mouse proteins as pheromone receptors. They have not yet carried out a key test to show that activating or eliminating the receptors alters a mouse's behaviour. But Buck says that her team is "intrigued by the possibility" of their being pheromone receptors, and is embarking on tests to find out.

So far, most evidence for human pheromones has come from behavioural research. In one study, scientists at the University of Chicago showed that women's menstrual cycles changed in length after they sniffed the sweat on pads previously worn in the armpits of another woman (K. Stern & M. K. McClintock *Nature* **392**, 177–179; 1998). Another study showed that newborn babies move towards a pad carrying the smell of their mother's breast rather than a clean pad (H. Varendi & R. H. Porter *Acta Paediatr.* **90**, 372–375; 2001).

But many researchers remain sceptical about the power of human pheromones — or even whether they exist at all. Human behaviour is influenced by so many conscious and unconscious factors that it's hard to measure the subtle changes in behaviour caused by a putative pheromone. "Work on human pheromones seems always to capture both public interest and scientific scepticism," says Holy.

If the receptors discovered in mice are indeed pheromone receptors, this could allow researchers to ask concrete questions about human pheromones. Scientists could, for instance, study people with mutant versions of the receptors to see if they behave differently when exposed to candidate pheromones.

The latest study "makes it harder for the sceptics," says Martha McClintock, who led the University of Chicago experiments. "Hopefully, this will catalyse a lot of research by high-powered scientists."

Neuroscientists already know that genes in neurons in the lining of the nose make more than 1,000 different olfactory receptors for identifying chemical odours. The discovery of this gene family won Buck and Richard Axel the 2004 Nobel Prize in Physiology or Medicine.

To look for other classes of receptor, Liberles and Buck isolated neurons from mouse noses and searched for genes that were active only in those cells and were similar to olfactory receptors. They hit on a family of 15 such receptors, called trace amine-associated receptors, or TAARs. There are six genes that code for similar receptors in humans.

The researchers next searched through a set of other chemicals to find those that activated the TAAR receptors. Three of the chemicals they found are present in mouse urine, which is known to affect the animals' social behaviour. One in particular earns the label of a pheromone because it is produced in the urine of male mice and accelerates the onset of puberty in females.

Experts say there may be other as yet undiscovered sensory receptors involved in detecting mammalian pheromones, perhaps including human ones. A paper in 2000 showed that a gene for one mouse pheromone receptor in the vomeronasal organ is also active in the human nose

(I. Rodriguez *et al. Nature Genet.* **26**, 18–19; 2000). "There might be other unidentified ones out there," says neurobiologist Frank Zufall of the University of Maryland in Baltimore.

Liberles and Buck are now testing whether there are molecules in human urine, sweat and vaginal fluid that can trigger the receptors and that might be human pheromones. They also want to examine whether the neurons manufacturing the newfound receptors are wired up in the brain in a different way from regular, smell-detecting neurons. If so, they might connect to centres that control our behaviour or emotions.

**Helen Pearson**

**"The new study makes it harder for the sceptics."**





D. BERHULAK/GETTY

Farmers in Niger must plant when the rains come — but will they last?

# Meteorologists pour into west Africa

The equipment list is impressive: several hundred researchers, five satellites and six research aircraft, including a converted Russian spy plane. In an unprecedented study of African weather patterns, atmospheric scientists from around the world have decamped this summer to the west of the continent. Their goal is to deliver the first truly useful forecasts of the region's violent monsoon storms.

If successful, the €50-million (US\$64-million) African Monsoon Multidisciplinary Analysis (AMMA) project will spawn hundreds of research papers and reshape thinking on west Africa's climate. But it should also help west Africans to cope with the destabilizing effects of unpredictable storms.

The monsoon rains not only dictate agricultural success or failure, but also bring floods that can damage roads and bridges, and encourage malaria-carrying mosquitoes. If these storms could be predicted weeks or months in advance, people in the region could plan ahead with far greater confidence.

"To have that would be a huge help," says Lauren Gelfand, who works on development projects for the charity Oxfam in Dakar, Senegal. "This is one of the driest regions in the world. People rely on one season of rain."

To develop their forecasting tools, AMMA has first had to plug holes in the region's meteorological networks (see *Nature* 435, 862–863; 2005). "There are precious few observations over the Sahara," says Doug Parker, an atmospheric physicist at the University of Leeds, who heads the British AMMA team.

European funding agencies have already spent around €2.5 million to upgrade west Africa's weather-balloon capability, particularly in five of the countries most involved in the project: Niger, Benin, Burkina Faso, Mali and Senegal. But this August's push is the heaviest data-gathering period planned in the roughly decade-long AMMA project, which began in 2002.

Last week, Parker said his team were still recovering from a series of bumpy four-hour data-collection flights around west African storms, which are some of the most intense on the planet. His group will team up with four other aircraft to fly coordinated missions until 21 August. British and French teams plan to drop curtains of radiosondes — meteorological devices that record and transmit data on temperature and pressure — in front of and behind storms. German and French teams using converted executive jets may fly right through the middle of the storms, at heights of around 15 kilometres. A former spy plane operated by Russian researchers will monitor the storms from above, at heights of around 25 kilometres.

Researchers hope that the flights will help them to understand the generation and movement of thunderstorms. The data gathered should also allow modellers to refine their predictions of everything from daily weather forecasts to long-term climate change.

But the most useful result for local people

will be more accurate rainfall forecasts, says Thierry Lebel, a member of the French AMMA team at the Grenoble Institute of Technology. At the start of the monsoon, for example, early rainfall can be followed by weeks of drought that can kill young plants, as Gelfand says has happened this year. If such events could be predicted, farmers could be warned to delay planting.

Results from AMMA could also help much farther away. Some 550 kilometres west of Senegal, a US team based in the Cape Verde

Islands in the Atlantic Ocean will complement the European operations. The researchers will make 10 to 12 flights on a DC-8 aircraft owned by NASA, gathering data on cloud particles, wind speed and direction,

rainfall rates and more. They hope to help understand how the Atlantic hurricanes that strike North America form, something that one team member, hurricane consultant Jeff Halverson, formerly of NASA's Goddard Space Flight Center, calls "one of the great unsolved mysteries of atmospheric science".

Monsoon storms can develop into hurricanes off the west African coast, if the right factors are present. But not all do: even given those factors only 10% of storms become hurricanes, and researchers struggle to predict which ones could go on to devastate the US east coast.

**Jim Giles**

See Editorial, page 485.



# Home health tests are 'genetic horoscopes'

## WASHINGTON DC

Picture this: two people, eager to prevent future ailments, order genetic tests over the Internet. The results are sobering — risks of osteoporosis, heart disease, diabetes and more. Affiliated companies then offer nutritional supplements to stave off these predicted sicknesses — but the pills turn out to be little more than multivitamins, offered with a hefty dose of misleading medical advice.

Fortunately, these two people are fictitious, part of an investigation conducted by the US Government Accountability Office (GAO). The inquiry found that 'nutrigenetic' tests, combining diet advice with genetics testing, are at best ambiguous and at worst dangerous.

On 27 July, the Senate special committee on ageing conducted a very public grilling of company representatives and federal regulators. Participants heard that the GAO had surreptitiously tested four companies that offer

home genetic testing to consumers. Three of the four testified: Genelex, based in Seattle, Washington; Sciona of Boulder, Colorado; and Suracell of Montclair, New Jersey.

Investigators posed as 14 consumers but used the DNA from just two people, a 48-year-old man and a 9-month-old girl. Despite this, the test 'results' were contradictory and warned of risks for various conditions. "It's clear they went way out ahead of the science," says Alan Guttmacher, deputy director of the US National Human Genome Research Institute in Bethesda, Maryland.

Two of the companies disagree. Howard Coleman, chief executive of Genelex, claimed at the hearing that the investigation's results were made specious by the 14 differing 'lifestyle questionnaires' that the investigators submitted along with the DNA samples.

The cost of at-home genetic tests in the study ranged from \$89 to \$395. Some are tantamount to "genetic horoscopes", said Thomas Hamilton, director of the survey and certification group at the US government's Center for Medicaid and State Operations. Unlike other witnesses, however, he argued that more regulation is not needed.

The Food and Drug Administration (FDA) also monitors some such tests because they count as medical 'devices', says Steven Gutman, director of the agency's office for diagnostic devices. The agency is investigating whether some of the tests should have undergone FDA review, he says.

Advocates of testing point out that the dubious nutrigenetic tests are very different to legitimate, reliable at-home tests for diseases such as cystic fibrosis. ■

**Gene Russo**

**"It's clear they went way out ahead of the science."**

## Cassini finds Titan's lakes after two-year search

Before Cassini reached Saturn, astronomers were excited about the chance of finding lakes of liquid methane or ethane on the planet's moon Titan. It is certainly cold enough there, and they figured that the gaseous hydrocarbons in Titan's atmosphere must be coming from somewhere.

But when Cassini started sending back images after its first flyby in July 2004, apparent river channels and seabeds were dry. And the Huygens probe, which landed on Titan in January 2005, found only solid, if damp, ground.

Now astronomers have found what they were looking for. Cassini's 17th close flyby — the first to scan Titan's colder northern hemisphere — has at last confirmed the presence of giant lakes. Further images should flood in, with 28 more flybys scheduled up to June 2008.

## India and Pakistan further their nuclear ambitions

With India poised to receive nuclear aid from the United States, its rival Pakistan has been working on nuclear plans of its own. It seems to be building a massive reactor that could produce 50 nuclear weapons a year.

The construction was discovered in commercial satellite images and made public on 24 July by analysts at the Institute for Science and International Security, a Washington-based arms-control group. The group says that the reactor, which seems to be several years from completion, could enable Pakistan to mass-manufacture plutonium weapons.

Two days later, the US House of Representatives approved further nuclear aid to India by a vote of 359 to 68. Arms-control experts fear that the accord, expected to pass the Senate later this year, will allow India to use domestic nuclear plants to boost its own weapons production.



Nuclear plan: view of Pakistan's Khushab complex, which could produce weapons-grade plutonium.

## Ocean hosts a reserve of rare genes

There are 10–100 times more bacterial species lurking in the ocean than previously thought, say marine researchers. By sequencing small snippets of DNA, Mitchell Sogin of the Marine Biological Laboratory in Woods Hole, Massachusetts, and his colleagues have discovered more than 20,000 different kinds of microbes in a single litre of sea water.

The scarcest creatures form what the team calls the 'rare biosphere'.

Their role is unclear, but they hold a reserve of genetic information that could help them survive, or even become dominant, if environmental conditions change. The study is part of the ongoing International Census of Marine Microbes, a ten-year project that began in 2000.



B. MORRIS/MICRO-SCOPE

## At last, a bird-flu vaccine that works at low doses

In the event of a bird-flu pandemic, a vaccine that works at low doses would be critical, enabling high numbers of doses to be made using the world's existing vaccine-production capacity. Last week, GlaxoSmithKline announced promising clinical-trial results — two shots of a vaccine containing just 3.8 µg of antigen gave 80% of 400 test subjects a strong immune response to the virulent H5N1 flu strain.

Existing production capacity could in theory produce enough of such a vaccine to treat 1.8 billion people. In contrast, an H5N1 vaccine from the US National Institutes of Health, tested last summer, required two doses of 90 µg — so the same amount of vaccine would treat less than 100 million people (see *Nature* doi:10.1038/news050808-9; 2005).

The company has yet to publish supporting data on the trials.

## UK centre offers cheap IVF in return for research eggs

Britain has edged nearer to allowing women to be paid to donate eggs for research purposes. The move is unusual — the US National Academy of Sciences recommended last year that women donating eggs for research should be paid expenses only, partly to avoid financial inducement for the risky procedure.

On 27 July, the UK Human Fertilisation and Embryology Authority gave the Newcastle Fertility Centre at Life permission to charge women less for *in vitro* fertilization (IVF) if they agree to donate any excess eggs to the centre's research. Donors will pay half of the normal £2,500 (US\$4,700) IVF fee. Newcastle researchers say they started receiving enquiries as soon as news of the decision became public.

Alison Murdoch, the centre's director, thinks donors will supply between five and ten eggs per week. Her group, which has created one cloned human embryo, currently uses eggs left over from IVF treatments, and has fewer than ten eggs a month to work with.

## Conflicts-of-interest keep hitting US headlines

Conflict of interest is like pornography: most people feel they know it when they see it, but it's deceptively difficult to define. In the United States last week, there was angry debate over various cases.

Some people were outraged by reports that the University of Virginia's Pat Michaels, a climatologist who says global warming is not a problem, has received several hundred thousand dollars from coal-burning power companies. But to Michaels and his supporters it was an above-board consulting deal.

And the Center for Science in the Public Interest reported that 18% of scientists on committees at the National Academies had conflicts of interest. Critics argued that the centre selected the committees most likely to be conflicted, and used impossibly strict criteria.

Political momentum to combat scientific conflicts of interest does seem to be growing, however. The Food and Drug Administration said on 24 July, for example, that it will review its policy of giving waivers to members of its advisory panels who have conflicts of interest. This is probably a bid to pre-empt a move by Congress that would force them to give up waivers altogether.

### Correction

The News Feature "The trouble with replication" (*Nature* **442**, 344–347; 2006) incorrectly stated that Dominique Bonnet has derived pluripotent stem cells from human bone marrow. Bonnet has derived such cells from mouse bone marrow only.

## BUSINESS

# Gas for the greenhouse

For big oil companies, carbon dioxide is waste; for people who grow fruit, it's a valuable commodity. **Ned Stafford** reports on a marriage of convenience in the Netherlands.

Dutch greenhouses, which grow much of Europe's fresh fruit, vegetables and flowers, have a surprising side effect: they emit carbon dioxide. Thousands of tonnes of the gas are generated to swell the serried ranks of tomato plants, but the plants use only a fraction of it, leaving the rest to seep into the atmosphere. Now, an innovative project is working to curtail wasteful emissions.

The project takes advantage of the fact that oil refineries can generate almost-pure carbon dioxide as a waste product. The gas is normally released straight into the atmosphere, but last autumn the project began pumping it to greenhouses instead. Owners of these greenhouses are now able to switch off the heaters they normally use to generate carbon dioxide, which saves them money. And, together with the refinery, they will garner government credits for cutting overall carbon dioxide emissions.

The €100-million (US\$130-million) project, known as OCAP, is the brainchild of two Dutch engineers: Jacob Limbeek and his former boss Hans Tiemeijer, who died in 2004. It exploits the link between the surplus carbon dioxide at Shell's refinery in Pernis, outside Rotterdam, and the demand for the stuff at the massed banks of greenhouses that stretch from there to Amsterdam, 85 km up the road.

The project also exploits the fact that the Dutch government obligingly constructed a gas pipeline between the two cities, three decades ago. The pipeline was never used — until OCAP brought it into service last September, as the spine of its carbon dioxide distribution network.

The idea for using waste carbon dioxide has been doing the rounds for about ten years. It was first championed by Tiemeijer and Limbeek when they worked for Energie Delfland, a Delft-based energy company. But the company dropped the plan after being taken over. The two men left and hawked the idea around for two years before wheedling

**"We believed in the project and that kept us on our feet."**



Line of beauty: a network of pipes will transport carbon dioxide from oil refineries to greenhouses.

the necessary capital from construction firm VolkerWessels, based in Rotterdam, and Hoek Loos — an industrial gas supplier and subsidiary of German conglomerate Linde. "It was difficult," says Limbeek. "But we believed in the project and that kept us on our feet."

Peter Ripson, a director of Hoek Loos, says the investment was a risky one for his company, but was too good to pass up. He won't say how much revenue the project will generate in its first year, but observes: "If there had been no business case for OCAP, we would not have done it."

The project bought the unused gas pipeline from the Dutch government for an undisclosed sum, and added 8 km to connect it to the refinery. Four control stations were constructed along the main pipeline and 150 km of piping was laid to deliver the carbon dioxide directly to some 500 greenhouses, which cover 1,300 hectares of cultivation between them.

The greenhouses are part of the €8-billion horticultural business of the Netherlands. The owners know they can speed plant growth by up to 25% by doubling carbon dioxide levels inside the greenhouses from the usual 380 parts per million. They used to do this by burning natural gas in heaters — and still do so in winter, when the heat is needed to keep plants warm. But in the summer, the heaters are now switched off.

According to Shell, 95 million cubic metres of natural gas will be saved each year as a result, and carbon dioxide emissions will be reduced by 170,000 tonnes.

The project is supported by a €4-million innovation grant from the Dutch government and €14 million in tax relief that Hoek Loos

and VolkerWessels will receive under a government scheme to encourage companies to conserve energy. The system is capable of delivering 160 tonnes of carbon dioxide an hour during peak months, and has been working at maximum capacity this summer. It charges the greenhouse owners between €40 and €70 per tonne — little more than half what it would cost them to generate carbon dioxide with gas heaters, Limbeek says.

## Cash in hand

A Shell spokesman says that OCAP is paying it for the carbon dioxide. The oil company also expects to receive credits for the reduction of its emissions of carbon dioxide into the atmosphere, under a European Union scheme that encourages such cuts (see *Nature* 441, 405; 2006). From 2008 onwards, the company will share its credits with the greenhouses, says the spokesman.

Almost all of the refinery carbon dioxide that is pumped through the system does end up in the atmosphere, says Gerhard Kerstiens, a plant scientist at the University of Lancaster, UK, who specializes in how plants use carbon. But total emissions are lowered because the greenhouses aren't generating the gas for themselves.

The project is thought to be the first of its type in the world. But Hendrik-Jan van Telgen, a plant scientist at Wageningen University, says he thinks it could work elsewhere.

For Limbeek — who is still only 34 — the project's completion is something of a personal vindication. "It took a lot of effort to come all this way," he says. "We had to overcome a lot of difficulties. It makes me proud."





# WHAT CHEMISTS WANT TO KNOW

Chemistry is a key component in all the scientific disciplines. But does that mean it is nothing more than a handy tool — or are there still major chemical questions to crack? **Philip Ball** finds out.

Physicists do not shy away from promoting the big questions that drive their field — how the Universe began, say, or what governs the behaviour of space, time and matter over scales from the atomic to the cosmic. Biologists, too, are happy to point to Erwin Schrödinger's question 'What is life?', which they are attempting to answer by unravelling DNA and mapping out the structures and interactions of proteins.

But what of the third basic science in the curriculum? To judge from the scant attention chemistry gets in the public media, you could be forgiven for thinking that it is a discipline whose time has passed, its grand puzzles all now answered. Does chemistry have any big questions left?

Identifying such frontier questions seems all the more urgent because university chemistry departments are facing an uncertain future. The department at the University of Sussex in Britain, for many years home to Nobel laureate Harry Kroto, co-discoverer of buckminsterfullerene, was the latest in a long line

threatened with closure. It has so far resisted an attempt to remodel it as a 'chemical biology' adjunct of the life-sciences division; but several other UK chemistry departments have failed to evade the axe. Following similar moves and concerns in the United States, a 2004 editorial in *Chemical and Engineering News*, published by the American Chemical Society (ACS), proposed changing the organization's name, rebranding it as the Society for Molecular Sciences and Engineering.

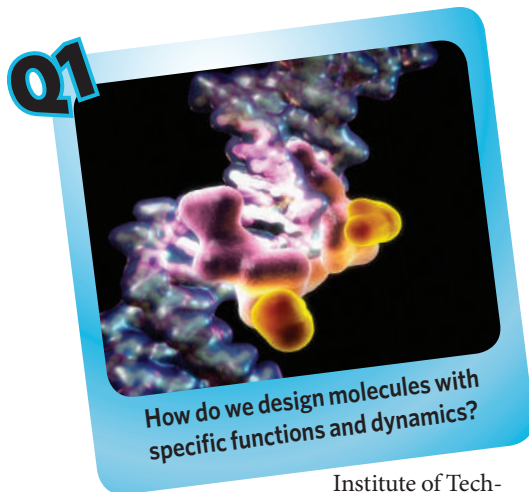
With departments closing and student numbers dwindling, can today's chemists be sure that their discipline will continue to be seen as a core science? Some of them complain that many of its most important questions are being framed in terms of the 'chemical' aspect of another discipline, rather than being seen as central to chemistry itself. In an attempt to gauge the prospects for academic chemistry, *Nature* asked many leading chemists what the field's big questions are, and whether in fact chemistry needs big questions to maintain a sense of coherence and identity.

The strongly synthetic character of chemistry sets it apart from the 'discovery' sciences such as physics, biology, astronomy and the Earth sciences. "Chemistry creates its object," as the French chemist Marcelin Berthelot wrote in 1860.

Many chemists still see this creativity as one of the field's strengths. "It makes chemistry able to set goals of a type most other sciences cannot hope to attain," says Ron Breslow, an organic chemist at Columbia University in New York and a past president of the ACS. "Where is synthetic astronomy — changing the gravitational constant to see what effect that has on the properties of the Universe, and thus perhaps improving it?"

And although synthetic biology is now emerging as a genuine discipline, to many chemists this is just another branch of applied chemistry, relying as it does on chemical techniques such as DNA synthesis and protein design. "We are the only science where things can be made that were never made before," says nucleic-acid chemist Jacqueline Barton at the California

J. MAGEE



**Q1**  
How do we design molecules with specific functions and dynamics?

Institute of Technology in Pasadena.

The downside of this focus on making stuff is that chemists can be portrayed as inveterate tinkerers — tweaking the molecular world to satisfy their curiosity, sometimes for fun and sometimes for profit. And it makes it especially hard to see where industrial chemistry ends and academic chemistry begins, because important practical challenges provide the motivation for much academic creativity.

“Chemistry is the scientific enterprise that fuels industry,” notes Barton. “Not just petrochemicals, but pharmaceuticals, biotechnology and computer chips.” Breslow agrees that chemistry faces not so much big questions as big practical challenges, such as “to devise a practical method to derive our needed energy from sunlight; to create a room-temperature superconductor that can carry large currents; to learn now to perform the manufacturing we need without damaging the environment”.

### Be specific

No one would deny the importance of applied and industrial chemistry. But if chemistry’s questions aren’t so much about what we can know but about what we can do, does that make it a form of engineering — a quest for particular solutions to particular problems?

According to inorganic chemist John Meurig Thomas of the Royal Institution in London, it is in the nature of chemistry to be a science of particulars. One can identify general principles of chemical bonding, for example, but what often matters is how these are enacted and modified in specific molecules. Similarly, he says, “it would be ludicrous to look for a general theory of catalysis that applies to all enzymes, materials, surfaces and so on.”

With so many chemists happily focused on practical goals, and with other disciplines nibbling away at the edges of chemistry, are there any big questions left at the academic core of the subject? And if there are, do they have the intellectual excitement of those

at the frontiers of physics and biology?

Chemists certainly have the tools and concepts to help answer some of the frontier questions arising in these other disciplines. The clearest consensus among the chemists approached by *Nature* was that many of chemistry’s most urgent questions are ones that address aspects of biology. “To me, the big unanswered questions concern the chemistry of life processes,” says physical chemist Richard Zare of Stanford University. Barton agrees: “A real understanding of biological processes always comes down to understanding the chemistry.”

### Essence of life

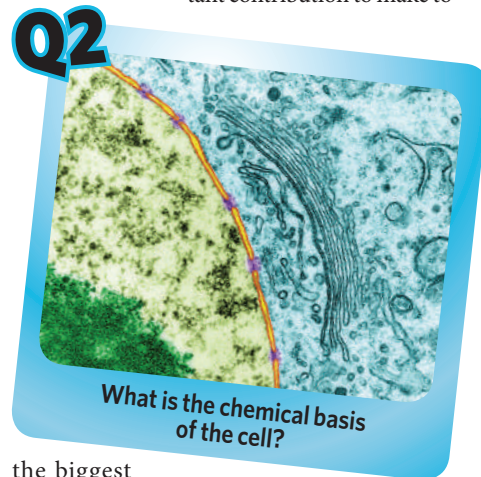
Harvard University chemist George Whitesides goes further. “The nature of the cell is an entirely molecular problem,” he says. “It has nothing to do with biology, really.” Whitesides suggests that the “really intellectual” parts of biology, such as its more quantitative and molecular aspects, are overlooked whenever biologists study whole organisms. These are strong claims, which biologists might contest. One can claim that the cell has nothing to do with biology, points out molecular biologist and Nobel laureate Sydney Brenner at the Salk Institute for Biological Studies in San Diego, California, but by the same token, one could say that all of chemistry is just quantum mechanics.

Still, many of the gaps in scientific understanding of the fundamental processes of molecular biology, such as protein folding, genetic encoding of biomolecular function, and highly selective molecular recognition, are fundamentally chemical problems. And although molecular biologists may assume that these things are broadly understood, from a chemical viewpoint they get more puzzling the closer one looks. Scientific understanding is still not good enough to provide a rational and predictive basis for the kind of molecular-scale interventions needed in biomedicine and drug development.

Meanwhile, the chemical nature of biomolecular processes such as signal

transduction, identified as a key question by chemical engineer Matthew Tirrell at the University of California, Santa Barbara, is one of the issues that is positioning chemistry as an information science. The concept of a lock and key to explain biomolecular recognition, proposed by German chemist Emil Fischer in 1894, can now be seen as the start of what supramolecular chemist and Nobel laureate Jean-Marie Lehn of the University Louis Pasteur in Strasbourg, France, has called the science of informed matter.

The concepts behind self-assembly have accustomed chemists to the idea that molecules can be programmed to interact and come together in very specific ways, and artificial replicating molecules have demonstrated the principles by which chemical information can be transmitted and amplified. “For me, chemistry has a most important contribution to make to



**Q2**  
What is the chemical basis of the cell?

the biggest question of all: how does self-organization arise and how does it lead the Universe to generate an entity that is able to reflect on its own origin?” Lehn says.

Lehn believes the next step is the design of chemical ‘learning systems’ that are not just programmed for assembly but can be trained. Indeed, another of the key questions in chemical biology raised by several chemists was the chemical basis of memory.

“Once we know the answer, maybe we can design new thoughts and memories, or even just learn how to retain old ones,” Barton suggests. Whitesides wants to know how to use chemistry to merge silicon electronics with grey matter. “How do I plug my computer into my brain?” he asks. That might sound like a matter for neuroscientists and electrical engineers — but as the signals between neurons are chemically mediated, this kind of interfacing demands command of a chemical language.

These are appealing directions for chemists to study, but do they qualify as truly chemical questions? To Whitesides, that couldn’t be more the case. “I take the view that most of what is interesting in science is now chemistry,” he says. He argues that even some of the key questions in a field as apparently remote from chemistry as astronomy, such as ‘How many Earth-like



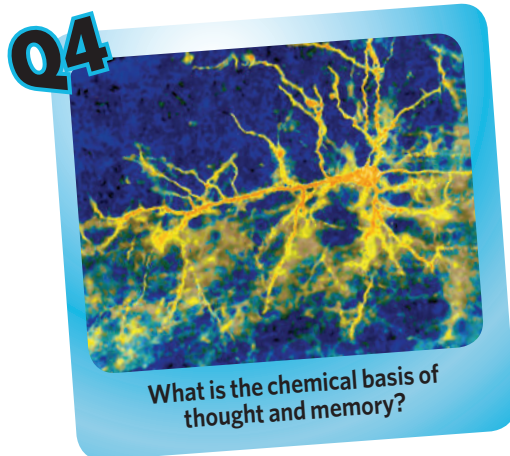
**Q3**  
How do we make the materials needed for the future, in energy, aerospace or medicine?



planets are there?' or 'What is on Saturn's moon Titan?' are fundamentally molecular ones.

When addressing interdisciplinary questions, what truly distinguishes chemists from physicists and biologists is that they are not content to ignore molecular-scale mechanisms. Tackling these issues confronts chemists with what is perhaps the core challenge of their discipline: to understand and predict the relationship between molecular structure and function.

Structure–property relationships are crucial to drug design, for example — one of chemistry's major concerns. "How do



**Q4** What is the chemical basis of thought and memory?

we encode specificity for particular cells, organelles or tissues into molecules?" asks Barton. "How do we make molecules go where we want them to go?" It is also essential for designing catalysts for use in industrial synthesis. But at present, a full understanding of the link between structure and function is often possible only for relatively simple, small molecules — and even then there are many details of the problem that have yet to be clarified.

### Action heroes

For example, Nobel laureate Ahmed Zewail, a physical chemist at the California Institute of Technology, points out that the dynamic behaviour of molecules can play as big a role in their reactivity as their molecular structure. It is now clear that the interactions between biomolecules aren't simply a matter of fitting a key into a lock — getting a good geometric match between the binding site and its target — but may depend on the dynamics of the interacting molecules and the solvent.

Chemists now think of reactions as happening on a complex, multidimensional energy surface, or landscape, which can be as rugged as a mountain range. So understanding protein folding is a question of how the molecule's peptide chain negotiates a trajectory across this energy landscape so that it ends up in the 'valley' corresponding to the correctly folded conformation. "In biology, thinking about the relationship between structure and function is not enough," says Thomas. "You've got to think about movement around the energy landscape." In other words the dynamics are key.

A big question for Zewail is how to control chemical function through dynamics. Compared with their ability to determine molecular structure, chemists have only just begun to understand what can be done to control reaction pathways in this energy landscape. In principle, this can be done by guiding molecules, perhaps using laser beams, into particular quantum states. So far this has been achieved for simple molecules, but looks extremely daunting for larger, 'floppier' ones.

And even if they crack the principles of molecular design, how do chemists apply them? "Until we reach the stage when someone can go in the lab and synthesize an arbitrary molecule in 100% yield in pure form without having a graduate student spend a year working it out, we have not really mastered synthesis," says Barton. "So the big question concerns the nature and rules governing how to assemble atoms into new molecules in a predictable and effective way. Then we could make whatever substances we want."

### Creative force

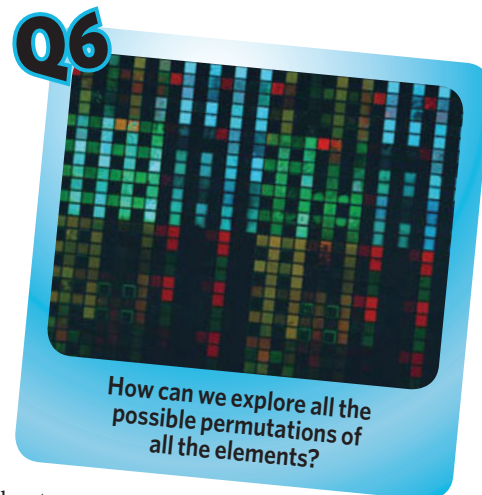
Only chemists know how truly difficult it is to engineer atoms and molecules — something that many other scientific disciplines rely on. If room-temperature superconductors or synthetic bacteria are ever created, it will not be physicists and biologists who make them. And if chemistry is chopped up and parcelled off to other disciplines, there will be no training ground for those who achieve such mastery over matter.

It would be wrong, moreover, to suggest that the heart of chemistry — rational synthesis — lacks intellectual appeal. Some argue that, rather than trying to understand the world, chemists are attempting to understand all possible worlds. "Chemistry has a useful aspect, but that is not the basic science," says Breslow. "The basic science is clear once we realize that the limited examples of molecules and reactions that nature has supplied are a microdrop in an enormous bucket compared with the wonderful chemical world still to be created and examined."

It has been estimated that there are



**Q5** How did life on Earth begin, and how and where might it begin on other worlds?



**Q6** How can we explore all the possible permutations of all the elements?

about  $10^{40}$  possible molecules that could be made from common elements with a molecular size comparable to that of a typical drug. "The known chemical world, including the expansion of the natural world that chemists have achieved, is nowhere near 1% of that," says Breslow.

It is this profusion that defies attempts to reduce chemistry to a handful of objectives. "Its universe is defined not by reduction to a few elementary particles, or even the hundred or so elements," says theoretical chemist and Nobel laureate Roald Hoffmann of Cornell University in Ithaca, New York, "but by reaching out to the infinities of molecules that can be synthesized. There is no end to the range of structure and function that molecules exhibit."

Most chemists seem content to work without big frontier questions to guide them. Such questions can be helpful to a discipline's sense of identity and direction, but they risk narrowing the possibilities of an inherently creative discipline. Some might argue that an excessive focus (at least in the public eye) on 'theories of everything' or decoding the human genome has not been terribly productive for physics or biology.

Besides, in chemistry, as in any science, the biggest breakthroughs often come from unexpected directions. "I do not think that I have ever identified in advance any of the major directions or key questions of chemical research that I witnessed in the past half century," confesses inorganic chemist Hubert Schmidbaur at the Technical University of Munich in Germany. "And it seems to me that the situation will be not very different in the next five decades."

"There is no Holy Grail in chemistry," Hoffmann admits happily. "Occasionally some are held up for public view," he says, but they are just "gimmicky candidates for the chalice". He adds that in a fundamentally creative field, the satisfaction comes from the chase, not the catch. "My natural philosophical disposition is not to work on big questions," says Hoffmann. "I like working on many detailed small problems in this wonderful chemical garden, while keeping my eyes open for the connections." ■

Philip Ball is a consultant editor for *Nature*.

For more on this topic, see Editorial, page 486.

PHOTOTAKE/ALAMY

NASA/JPL/SPACE SCI. INST.

H. KOINUMA & I. TEKEUCHI/NATURE MATER. 2, 429–438 (2004)



# CONSERVATION AT A DISTANCE

There's more to ecology than ringing birds, and in this special section *Nature* explores how the molecular sciences are transforming the field. In 'Atomic detectives', **Sharon Levy** explores how atoms in feathers can reveal the secrets of rare warblers. And **Carina Dennis** unveils a technique that aims to make killing whales for science a thing of the past in 'A gentle way to age'.

## Atomic detectives

In a scrubby jack-pine thicket in Michigan, Peter Marra and his colleagues set up a net, bait it with a plastic lemon, and play the song of one of the planet's rarest birds, the Kirtland's warbler. Soon after, a male warbler, resplendent in his bright yellow and grey breeding plumage, comes bombing into the net, mistaking the plastic lemon for a competitor on his turf. To reach this place, the tiny bird may have flown 2,400 kilometres from the island of Eleuthera in the Bahamas.

The Kirtland's warbler *Dendroica kirtlandii*, (pictured below and right) belongs to a family of small, colourful birds that nest in the forests of the United States and Canada, and winter far to the south. In the 1970s and 80s, destruction of its breeding habitat brought the species to the brink of extinction. During this period, observers found fewer than 200 singing males in Michigan. Intensive management of their breeding habitat — replanting jack pines and removing cowbirds, which parasitize warbler nests — has helped the Kirtland's population rebound to more than 4,000 birds.

But events in their largely unknown wintering grounds may still be harming warblers. Marra, a biologist at the Smithsonian Migratory Bird Center in Washington DC, is among a pioneering group that is trying to identify threats to warblers and other vulnerable creatures<sup>1</sup> — using not just bird nets and binoculars, but also mass spectrometers.

The tool relies on the fact that elements such as carbon, hydrogen and nitrogen have more than one isotope. The isotopes differ in their mass, so a mass spectrometer can reveal how much of each isotope a sample contains. Because different foods have different isotopic make-ups, the ratios of carbon and nitrogen isotopes in the tissues of animals can reveal what they have been eating, and suggest the type of area they've inhabited. And thanks to the fact that some elements, such as hydrogen, have isotope ratios that vary over the globe, researchers can also see where an animal has been, by tracing its path across the isotopic landscape, or 'isospace' (see 'Written in the elements').



Where to? Peter Marra is using isotopes to work out where the Kirtland's warbler (below) spends winter.

Such studies are allowing researchers to reconstruct the behaviour of many endangered species, yielding information that could prove invaluable to efforts to conserve them.

Many songbirds that make long annual migrations between northern latitudes and the tropics have been declining for decades. The reasons have confounded researchers, who until recently have not been able to connect breeding populations in the north with specific wintering populations in the south. "The

links between breeding and wintering grounds represent a huge hole in our knowledge," says Marra. "It doesn't matter how much conservation work we do in the United States. We need to protect animals year-round, in both breeding and wintering areas."

Marra's earlier study<sup>2</sup> of the American redstart (*Setophaga ruticilla*), a common warbler, was the first to show that habitat conditions in distant tropical wintering grounds could influence the birds' reproductive success in US forests. He exploited the fact that plants in cool, moist habitats tend to use a different form of photosynthesis from that used by most dryland plants. This means that plants in damp habitats have lower levels of heavy carbon (carbon-13) in their tissues. The different carbon ratios transfer up the food chain to the insects on which the redstarts feed.

Using the ratio of carbon-13 to carbon-12 in the birds' blood, Marra inferred that the healthiest males arriving at his study site in New Hampshire had wintered in damp mangrove forests, where they had fattened up on abundant insects. Because these birds had



Some migrants are at risk in their winter habitat.

C. I. BOCETTI

P. DORAN

become fit to take their long journey north sooner than birds wintering in the drier scrublands of Jamaica and elsewhere in the Caribbean, they arrived first on their northern breeding grounds, laid claim to the choicest nesting territories, and fledged more chicks.

Another ground-breaking study<sup>3</sup> used isotopes to reveal the wintering grounds of the black-throated blue warblers (*Dendroica caerulescens*). These warblers breed in forests throughout the eastern United States and Canada, but over the past 30 years or so the population in southern states has dwindled while numbers in the north have held steady. Dustin Rubenstein of Cornell University in Ithaca, New York, and his colleagues analysed hydrogen isotope ratios, which vary predictably in rainfall at different latitudes. They showed that birds breeding in the southern United States wintered on the island of Hispaniola, which includes Haiti, a country that has suffered drastic deforestation. Those breeding farther north wintered in more intact habitats in Jamaica and Cuba.

### What counts

"In North America, we're very good at running around counting birds, and knowing whether populations are going up or down," says Keith Hobson, a research biologist for the Canadian Wildlife Service who is based in Saskatchewan. "The problem is we don't often know why these changes are happening. So the insights that stable-isotope studies give us into causes are priceless."

The stable-isotope technique itself is not new; archaeologists have long used isotope ratios in bone to reconstruct the diets of ancient peoples. Forensic scientists are catching on to their use too (see 'Written in the elements'). But the high cost and relatively low efficiency of mass spectrometers meant that wildlife biologists were slow to take up the tool. This all started to change in the 1990s, when improved equipment became available. "There's been a great improvement in the speed with which samples can be run," says Hobson. "People can suddenly afford it, and the throughput of samples has become huge with continuous-flow isotope mass spectrometry."

Recent work reveals the changes that can reverberate throughout whole ecosystems when an animal population is decimated. In Monterey Bay, California, numbers of sardines and anchovies began to crash in the 1940s — possibly reflecting a natural oscillation exacerbated by decades of overfishing. This year, Ben Becker of the Point Reyes National Seashore in California and Steven Beissinger of the University of California, Berkeley, published a study in which they compared isotope ratios in the feathers of endangered modern marbled murrelets (*Brachyramphus marmoratus*) to those from museum specimens collected before the sardine crash<sup>4</sup>. Because the heavier isotope of nitrogen, nitrogen-15, concentrates in organ-

isms' tissues as it moves up food chains, the isotope ratio offers a guide to what position the animal occupies in the chain. Becker and Beissinger found that today's seabirds are forced to feed lower on the food chain than their ancestors, surviving on krill rather than more nutritious small fish. This may explain why, in some years, as little as 10% of local marbled murrelets attempt to nest.

### In at the krill

Steven Emslie of the University of North Carolina in Wilmington and Bill Patterson of the University of Saskatchewan in Saskatoon are also using isotopes to look at krill. They are asking whether penguins shifted to eating krill after whalers wiped out most of the Southern Ocean whales — major consumers of the crustaceans — some 200 years ago.

Hobson first applied isotope techniques to wildlife in the 1980s, when he used nitrogen isotope ratios to gather information on the feeding patterns of seabirds, including the long-extinct great auk (*Pinguinus impennis*)<sup>5</sup>. He now thinks the technique has most potential as a probe into the mysteries of migration. He and his colleagues are using carbon and hydrogen isotopes to reveal the surprising long-distance travels of dragonflies and bats, and to probe the fates of some once-abundant North American migratory waterfowl. The technique could even help answer the long-standing puzzle

**"The exciting thing is that the isotope approach can be entirely non-destructive."**  
— Keith Hobson

## WRITTEN IN THE ELEMENTS

The dead speak, as any forensics expert will tell you. Even the very atoms from which a body is made can have something to say about where the dead person was from, how old he or she was, or even where the person had recently travelled.

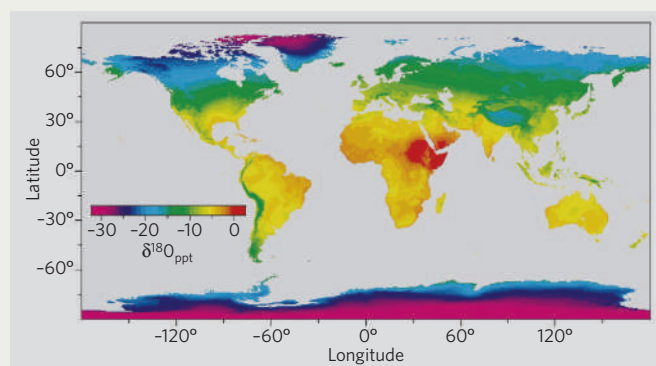
The technique, called stable-isotope analysis, relies on the fact that many elements come in heavier and lighter forms, or isotopes. The ratios of these isotopes varies in space and time, and this helps scientists to coax information from corpses, as well as from bioterror microbes and drug hauls.

Wolfram Meier-Augenstein, for example, a chemist at Queen's University in Belfast, UK, has assisted in six murder investigations, determining a body's area of origin from its isotope ratios. The amount of oxygen-18, for example, varies in water sources across the globe<sup>7</sup> (see map). And sugar in the United States is more likely to be

made from maize, which has more carbon-13 than other sources used in Europe, such as sugar beets. "Quite literally, you are what you eat and drink," says Meier-Augenstein.

At the Lawrence Livermore National Laboratory in California, Bruce Buchholz can tell the birth date of a person to within a year and a half from a single tooth by taking advantage of the spike in carbon-14 that occurred on Earth in the late 1950s due to atmospheric testing of nuclear weapons<sup>8</sup>. The technique was used to help identify some victims of the December 2004 Sumatran tsunami.

Drug dealers and bioterrorists are also a target. Jim Ehleringer at the University of Utah in Salt Lake City has used isotope-ratio signatures to determine the origin of cocaine. Bolivian coke, for example, has less nitrogen-15 than coke from Peru<sup>9</sup>. He and Helen Kreuzer-Martin, now at the Pacific Northwest National



Laboratory in Richland, Washington, have been gathering the data needed to assign geographical origins to dangerous microbes, such as anthrax or salmonella, and their spores<sup>10</sup>. Stable-isotope analysis is ready to go mainstream, says Kreuzer. "My guess is that we will have it in the courtroom in the next few years or so."

"The technique is probably in wider use than is generally known," says Meier-Augenstein, who adds that law enforcers

would prefer to keep it under their hats, so criminals are at a disadvantage. The FBI has its own expanding stable-isotope lab, and some commercial labs exist.

Meanwhile, Meier-Augenstein is looking at parts of the body that record isotope ratios on relatively recent timescales. "It is very good for verifying a person's movements," he says. "I have people knocking on my door from anti-terrorist forces asking what I can do for them."

**Emma Maris**

SOURCE: REF. 7





T. DAVIES/CORBIS

of how new travel routes evolve.

Stuart Bearhop of Queen's University in Belfast, UK, studied hydrogen isotope ratios in the claws of a community of European blackcaps (*Sylvia atricapilla*). All European blackcaps breed in Germany and Austria, but this population diverged from its parent population, which winters in Spain and Portugal, and began wintering in Britain a few decades ago<sup>6</sup>. Birds that winter in Iberia have significantly higher levels of hydrogen-2, allowing researchers to tell the two groups apart even though they mix on their breeding grounds. Blackcap males from Britain arrive at the nesting grounds earlier, tend to mate with females who have also come from the United Kingdom, and produce more chicks than birds that come from the south.

### The mechanics

The competitive advantage of British blackcaps almost certainly arises from a change in the timing of their travel to the nesting grounds. "Migratory behaviour in this species is triggered by changes in day length, and the critical day length is reached in Britain about ten days earlier than it is in Spain and Portugal," says Bearhop. The altered timing gives males the first choice of breeding territories, and helps ensure that UK-wintering birds find like-minded mates. Although there is as yet no evidence that the two populations are speciating, isotope ratios have made clear what Bearhop calls "a nice mechanism by which two populations could become reproductively isolated."

Even the hidden physiologies of animals can be revealed by isotope analysis. The heavier isotope of nitrogen occurs in much higher concentrations in marine than in terrestrial or



**Species across the globe, from penguins to marbled murrelets, could benefit from isotope tracking.**

freshwater food webs. Hobson is exploiting this pattern to study birds that winter on the ocean but fly far inland to breed. "The dogma has been that large-bodied, high-latitude breeding birds carry all their nutrients with them when they fly inland, that they use marine nutrients stored in fat to make their eggs," he says. This idea led to concerns that the chicks of many waterfowl species that breed in the high Arctic were suffering from heavy loads of contaminants, such as mercury and selenium, that pollute nearshore waters where the adult birds forage in winter.

But Hobson's studies of the feathers of newborn birds show that most marine ducks and geese create their eggs with nutrients from food they find on their nesting grounds. That finding relieves some concerns over marine contaminants, but also points out potential problems. "If these birds rely on local conditions to favour reproductive success," says Hobson, "then climate change in the north may have a much greater influence on them than expected."

Although they face the important challenge of improving and updating hydrogen-isotope

maps — across latitudes and elevations — stable-isotope techniques have opened up a world of possibilities for wildlife researchers. Isotope ratios in different types of tissue offer a way to analyse food sources over varying periods of time: ratios in faeces or blood plasma reflect food eaten over a day or two; those in cellular blood cover diet over about a month. And the feathers or claws formed on breeding grounds

retain the same isotope signature many weeks later when birds arrive to winter in an entirely different part of the globe.

All this represents a huge improvement on traditional methods: hoping that birds banded near their nests will be sighted on distant, unknown wintering turf, or killing a research subject to get a snapshot of its stomach contents. The ultimate beauty of stable-isotope studies is that they allow information on whole populations of animals to be collected without injuring a single individual. "The exciting thing is that the isotope approach can be entirely non-destructive," says Hobson. ■

**Sharon Levy is a science writer based in California.**

1. Hobson, K. A. *Oecologia* **120**, 314–326 (1999).
2. Marra, P., Hobson, K. A. & Holmes, R. T. *Science* **282**, 1884–1886 (1998).
3. Rubenstein, D. R. *et al. Science* **295**, 1062–1065 (2002).
4. Becker, B. H. & Beissinger, S. R. *Conserv. Biol.* **20**, 470–479 (2006).
5. Hobson, K. A., Piatt, J. F. & Pitochelli, J. J. *Anim. Ecol.* **63**, 786–798 (1994).
6. Bearhop, S. *et al. Science* **310**, 502–504 (2005).
7. Bowen, G. J. & Wilkinson, B. *Geology* **30**, 315–318 (2002).
7. Spalding, K. L., Buchholz, B. A., Bergman, L.-E., Druid, H. & Frisén, J. *Nature* **437**, 333–334 (2005).
8. Ehleringer, J. R., Casale, J. F., Lott, M. J. & Ford, V. L. *Nature* **408**, 311–312 (2000).
9. Kreuzer-Martin, H. W., Lott, M. J., Dorigan, J. & Ehleringer, J. R. *Proc. Natl Acad. Sci USA* **100**, 815–819 (2003).

J. P. BLAIR/NATL GEOGR. IMAGE COLLECTION





## A gentle way to age

The day started much like any other. Researcher Daniel Burns and his colleagues were keeping a respectful distance behind a pair of courting humpback whales. Then a third whale made an unexpected appearance. The new arrival began to press his attentions on the increasingly harassed female. Taking fright, she manoeuvred to place the research boat between herself and her suitors, leaving Burns and his team sandwiched between three 35-tonne whales, all shoving in different directions. “The adrenaline was pumping,” says Burns, recalling how he braced himself on the bow of the six-metre vessel as it lurched precariously between the massive animals, each measuring about twice the length of the boat. But he and his team got what they came for: a sample of whale ‘dandruff’.

Using a kitchen sieve strapped to the end of a stick, Burns and his team trail behind whales and quickly scoop up flakes of skin — some as big as the palm of a hand — before they sink. The flakes are naturally shed by the animals when they launch themselves out of the water (see above) or slap their tail fins on the surface.

Scooping up skin might not have the glamour normally associated with whale watching, but Burns and his colleagues think the flakes could offer a non-invasive way of working out a whale’s age. If they are right, one of the key arguments in favour of killing whales for scientific purposes will be dead in the water.

Burns, a PhD student at the Southern Cross University Whale Research Centre in Lismore, New South Wales, goes out in virtually all

weather, from dawn until dusk. He is part of a research team, led by Peter Harrison, that is developing genetic techniques to age humpback whales (*Megaptera novaeangliae*) from their sloughed skin. Currently, the most accurate way to age baleen whales such as humpbacks, which lack the teeth often used to age whales, is to count the lamination rings that form in their ear wax. Because the ear canal is closed to the outside world, wax accumulates in layers. Such laminations are conventionally thought to form twice a year in humpbacks. But counting them requires killing and dissecting the whale.

**“It’s better to have the approximate age of a live whale than the exact age of a dead whale.”**  
— Peter Harrison

### Showing the years

“Being able to age whales is something of a holy grail,” says Phil Clapham, a whale biologist at the National Marine Mammal Laboratory in Seattle, Washington. Knowing whales’ ages would enable biologists to model population dynamics, to better understand whale behaviour and assess how well the creatures are recovering from the devastating slaughter of the past century. It could also reveal whether mating tactics change with age, why individuals associate with each other at certain stages, and answer basic questions, such as exactly how long humpback whales live. Earlier reports

suggested half a century, but researchers now suspect they could live for twice as long as that.

Harrison’s team — and other groups around the world — hope that clues to a whale’s age lie in the tips of its chromosomes, where there are stretches of short, repeated DNA sequences called telomeres. Like countdown timers, telomeres gradually shorten with age in some species, including humans. Studies of humans and birds suggest that individuals can be assigned to broad age classes<sup>1,2</sup>, based on the length of their telomeres — although variability between individuals and in the rate of telomere loss during an individual’s lifetime make it difficult to determine age precisely. Whale researchers are optimistic that, if whales are also shown to lose telomere sequences with age, telomere analysis could provide a non-lethal means to assign individuals to broad age groups.

The need is urgent, say researchers, because last year Japan declared that it will double its catch of minke whales for scientific research. It will also include, for the first time since the International Whaling Commission (IWC) imposed a moratorium on commercial whaling in 1986, humpback and fin whales in its annual catch for research — with a yearly take of 50 of each species<sup>3</sup>. One of Japan’s main scientific arguments for increasing its catch is the need to collect samples to determine population structure. Harrison’s team and others hope that the telomere method could scotch this claim. “If the technique works, it would put a large nail in the coffin of Japan’s argument for a



scientific whaling programme,” says Clapham.

Using telomeres to estimate the age of whales — and other cetaceans, such as dolphins and porpoises — is untested water. Telomeres shorten with age in some animals, but not in others<sup>4</sup>; some even lengthen<sup>5</sup>. To show that the telomere ageing method will work in whales, researchers need a sample of known age.

Harrison's team is fortunate to have access to an extensive photo library of whales created by Trish Franklin, a historian, and her husband Wally, a retired airline executive. The couple have been meticulously photographing and documenting the animals since 1989. “It started off as an interest, which then became a passion — and later an obsession,” says Wally, who, along with his wife, is now undertaking a PhD under Harrison's supervision.

The photo library, comprised of pictures of nearly 3,000 individual whales, is invaluable as it provides the minimum age of many animals. The researchers are now systematically genotyping each animal — that is, obtaining an individual genetic ‘fingerprint’ — using the DNA extracted from sloughed skin samples.

### Signs of success

While Burns braves unpredictable waters to collect the skin samples, his colleague Martin Elphinstone processes their telomeres in the quiet calm of the lab environment. Some early results are encouraging. According to Harrison, preliminary data from a handful of humpback whales hint that calves and adults can be distinguished on the basis of telomere length. “Ideally, in the long term, we'd like to refine it to a 5- to 10-year band width,” says Harrison. Although unlikely to be as accurate as counting ear wax laminations, this range would be sufficient for population modelling, he says. “It's better to have the approximate age of a live whale than the exact age of a dead whale.”

Support for these data comes from Per Palsbøll, a population geneticist at the University of California, Berkeley, who is also studying telomeres in humpback whales. His team has examined the DNA in skin biopsies taken from a dozen whales, ranging from calves to 20-year-old animals. “From this tiny data set, we see longer telomeres in younger animals,” says Palsbøll. But he cautions that the method is far from reliable at this stage. “We have tremendous problems with reproducibility, even within samples from the same individuals,” he says.

Telomere work is a complicated business. To assess how the rate of telomere loss changes over an individual's lifetime, Harrison's team is testing multiple samples from the same individual at different times in its life. To do this, the group will tap into a treasure-trove of skin biopsies collected over several decades by Scott Baker, a molecular ecologist at the University of Auckland in New Zealand.

Baker has samples from northern-hemisphere humpback whales, ranging from 2 to 30 years old, which were resampled 14 years after the initial biopsies were taken. Harrison is



Daniel Burns (left) and his colleagues collect shed skin for DNA analysis by trailing after whales with a sieve.

collaborating with Baker to measure telomere length in these animals to see how it varies within and between individuals. Baker also has a collection of biopsies taken ten years ago from southern right whales (*Eubalaena australis*) from the Auckland Islands; working with Nick Gales, of the Australian Antarctic Division of the Commonwealth Scientific and Industrial Research Organization, based in Canberra, he is resampling the same animals for telomere analysis.

Harrison's method for analysing whale telomeres from sloughed skin samples is unique in that it is non-invasive; getting DNA from skin biopsies involves taking a plug of tissue from near the dorsal fin using a biopsy dart from a cross-bow or modified rifle.

But using whale skin flakes is not without problems. The skin is often damaged and dying tissue. “Many samples don't provide enough high-quality DNA,” says Elphinstone. It is also difficult to target specific whales using this technique, as skin flakes from different whales can get mixed up in the water. Harrison thinks this problem will be overcome once they have got a DNA fingerprint, or genotype, for every animal in their photo library to act as a reference.

If telomere analysis can be validated in humpbacks, it could be extended to other

whale species. Gales is examining telomeres in animals ranging from those that are relatively short-lived, such as harbour porpoises that live less than 20 years, to bowhead whales that may live for more than 200 years. “We are looking at animals of quite different lifespans to see whether telomere length correlates with longevity,” says Gales.

Gales is also analysing the method in a resident bottlenose dolphin population in Sarasota Bay, Florida — one of the world's best-studied dolphin populations. “The advantage is that a tremendous amount is known about their age and population structure, so telomeres can really be put to the test,” says Gales, who is collaborating on the project with Randall Wells at the Mote Marine Laboratory in Sarasota.

The developments can't come soon enough for the Australian researchers. Japan plans to catch humpback whales in the Antarctic feeding grounds in the south-latitude summer of 2007–08, which includes the population that Harrison's team is studying. They dread the thought that one of their whales, many of which have been given nicknames, will turn up on a whaling vessel.

Even if whale researchers can develop a non-lethal method to age the creatures, whether it would immediately halt scientific whaling is questionable. Japan's Institute of Cetacean Research declined to comment when approached by *Nature*. Gales, who heads the Australian delegation for the scientific committee of the IWC, thinks it wouldn't. “But at least we can use it to apply more political pressure,” he says.

**Carina Dennis is *Nature's* Australasian correspondent.**



Filter feeders: baleen whales don't have teeth, which are often used to age whales.

1. Tsuji, A., Ishiko, A., Takasaki, T. & Ikeda, N. *Forensic Sci. Int.* **126**, 197–199 (2002).
2. Haussmann, M. F., Vleck, C. M. & Nisbet, I. C. T. *Exp. Gerontol.* **38**, 787–789 (2003).
3. Gales, N. J., Kasuya, T., Clapham, P. J. & Brownell, R. L. Jr *Nature* **435**, 883–884 (2005).
4. Monaghan, P. & Haussmann, M. F. *Trends Ecol. Evol.* **21**, 47–53 (2006).
5. Haussmann, M. F. et al. *Proc. R. Soc. B* **270**, 1387–1392 (2003).

## The gender debate: science promises an honest investigation of the world

SIR — Ben Barres's Commentary article "Does gender matter?" (*Nature* **442**, 133–136; 2006) misrepresents my position.

In my book *The Blank Slate* (Allen Lane, London, 2002), and in a published debate ([www.edge.org/3rd\\_culture/debate05/debate05\\_index.html](http://www.edge.org/3rd_culture/debate05/debate05_index.html)), I reviewed a large empirical literature showing differences in mean and variance in the distributions of talents, temperaments and life priorities among men and women. Given these differences, some favouring men, some women, it is unlikely that the proportions of men and women in any profession would be identical, even without discrimination. That is probably one of several reasons that the sex ratio tips towards women in some scientific disciplines (such as my own, developmental psycholinguistics) and towards men in others. Barres renders this conclusion as "a whole group of people is innately wired to fail" — an egregious distortion.

Barres claims that I have denied that sex discrimination is a significant factor in professional life, whereas I have repeatedly stated the opposite, and indeed provided a jacket endorsement for Virginia Valian's book *Why So Slow?* (MIT Press, Cambridge, 1998) that summarized the evidence.

As for encouraging women in science: in my experience, students of both sexes are attracted to science because it promises an honest investigation into how the world works, an alternative to the subjectivity, simplistic dichotomies and moralistic name-calling that characterize politics and personal quarrels. Let's hope Barres's Commentary article does not discourage them.

**Steven Pinker**

Department of Psychology,  
Harvard University, Cambridge,  
Massachusetts 02138, USA

## Let's encourage gentler, more reflective scientists

SIR — I am one of three men that Ben Barres, in his Commentary article<sup>1</sup>, pins up in a rogues' gallery and accuses of crimes against women. His article invents and then criticizes the 'Larry Summers Hypothesis', for which he cites three sources<sup>2–4</sup>. I urge readers to look at those sources and ask themselves whether Barres has misrepresented our arguments and views.

Specifically, Barres accuses me of arguing that women have "lesser innate abilities": this is not true, it is not what I wrote, it is not what I meant and it is not what I

think. However, I do believe that there are significant differences in the psychological characteristics of populations of boys and girls as well as men and women (some differences favour females, some males) — a truth that Barres makes use of.

For example, if (as Barres states), men are responsible for 25 times as many murders as women, does that statistic make women inferior? I think not.

Likewise, if more women than men tend to avoid the vicious struggle to survive in science, is this an argument that women are less well equipped? Not so. In a recent article<sup>4</sup>, I argue that scientists need to open their doors to gentler, more reflective women and men. And that if we do, we will encourage talented but less aggressive people, and the proportion of women at the top in science will rise as a welcome side effect.

**Peter A. Lawrence**

Division of Cell Biology,  
Medical Research Council Laboratory  
of Molecular Biology,  
Hills Road, Cambridge CB2 2QH, UK

1. Barres, B. A. *Nature* **442**, 133–136 (2006).
2. Summers, L. Letter to the Faculty Regarding NBER Remarks [www.president.harvard.edu/speeches/summers/2005/facletter.html](http://www.president.harvard.edu/speeches/summers/2005/facletter.html) (2005).
3. *The Science of Gender and Science. Pinker vs. Spelke: A Debate* [www.edge.org/3rd\\_culture/debate05/debate05\\_index.html](http://www.edge.org/3rd_culture/debate05/debate05_index.html) (2005).
4. Lawrence, P. A. *PLoS Biol.* **4**, 13–15 (2006).

## Bias was built into research from the beginning

SIR — As a basic scientist attempting to understand the mechanistic basis of sex differences in the brain, I found Ben Barres's Commentary article<sup>1</sup> of particular interest.

The existence of robust and reliable sex differences in brain regions relevant to the control of reproduction cannot be refuted. However, my research group spent a year of frustration trying to apply the same principles to putative sex differences in cognition. This ended with the epiphany that even the standard laboratory rat shows few, if any, sex differences in the morphometry of regions relevant to cognition such as the hippocampus, and that the learning ability of both sexes is essentially the same<sup>2</sup>.

A search of the literature revealed that one of the first reports of sex differences or hormonal modulation of learning in rats, published in 1926, cites the 'fact' that such differences are reliably established in humans as supporting evidence (and note, the author was a woman)<sup>3</sup>. Thus, even the fundamental science of learning in animal models was tainted by bias from its inception. Only with the influx of a large number of women scientists is the notion of superior male spatial ability beginning to be challenged.

Data are now being reinterpreted as showing a difference in learning strategy rather than ability<sup>4,5</sup>.

When I had the fortune to meet Ben Barres a few years ago, he told me his anecdote about losing the ability to cry easily after changing his gender to male, and suggested that sex differences in the lachrymal gland would be an excellent topic for study. At the time, I laughed at his naive notion that a behavioural sex difference would be the result of a peripheral gland. But now, I am not so sure.

**Margaret M. McCarthy**

Departments of Physiology and Psychiatry,  
University of Maryland Baltimore School of  
Medicine, 655 W. Baltimore Street,  
Baltimore Maryland 21201, USA

1. Barres, B. A. *Nature* **442**, 133–136 (2006).
2. McCarthy, M. M. & Konkle, A. T. M. *Front. Neuroendocrin.* **26**, 85–102 (2005).
3. Ball, J. *Am. J. Physiol.* **78**, 533–536 (1926).
4. Perrot-Sinal, T. S., Heale, V. R., Ossenkopp, K. P. & Kavaliers, M. *Behav. Neurosci.* **110**, 1309–1320 (1996).
5. Beiko, J., Lander, R., Hampson, E., Boon, F. & Cain, D. P. *Behav. Brain Res.* **151**, 239–253 (2004).

## Holding the centre among the scatter-brained

SIR — In his Commentary article "Does gender matter?" (*Nature* **442**, 133–136; 2006), Ben Barres counters suggestions that women are underrepresented in the sciences because of innate ability.

Barres's graph of maths scores for boys and girls aged 4 to 18 shows that the differences in abilities are small. And yet most winners of the US Putnam mathematics competition each year are men. Putnam winners clearly fall on the far-right tail of the maths IQ curve, and it's precisely in the tails that the gender differences become clear. Male aptitude is more variable and, as Marilyn vos Savant points out in her excellent column "Are men smarter?" ([www.parade.com/articles/editions/2005/edition\\_07-17-2005/featured\\_0](http://www.parade.com/articles/editions/2005/edition_07-17-2005/featured_0)), men are overrepresented at both ends of the spectrum. (I call them 'scatter-brained'.)

Although I wouldn't hesitate to speak out in the face of blatant discrimination, as Barres urges, my advice differs from his. Grow a thick skin. If you encounter bias, choose your battles. Don't get distracted by the small stuff, but speak up when it really matters.

**Donna L. Dierker**

Department of Anatomy & Neurobiology,  
Washington University School of Medicine,  
660 South Euclid Avenue, St Louis,  
Missouri 63110, USA

Readers are encouraged to add their comments to the Ben Barres Commentary on the Nature News Blog at: [http://blogs.nature.com/news/blog/2006/07/does\\_gender\\_matter.html](http://blogs.nature.com/news/blog/2006/07/does_gender_matter.html)



## BOOKS &amp; ARTS

# Future perfect?

Dizzying advances predicted for the next century could improve the world — or lead to disaster.

## The Meaning of the 21st Century: A Vital Blueprint for Ensuring Our Future

by James Martin

Riverhead: 2006. 410 pp. \$26.95.

To be published by Eden Books in the UK in October.

### Norman Myers

The past 50 years have probably seen more change than the previous 500 years, and those 500 witnessed more advances than the previous 5,000. The current century will surely surpass the lot, with an onrush of scientific, technological, economic, social and political change, all of a character and on a scale way beyond anything we have experienced to date. The future will not be an extension of the past, but will mark a departure from much of our activities since the start of civilization 5,000 years ago.

We have recently had a plethora of books surveying humankind's niche on Earth — our numbers, our impacts, our advances and our setbacks. Many have examined our prospects for the future, but James Martin's *The Meaning of the 21st Century* is surely the most ambitious yet. He predicts a "grand transition" from "a planet on a self-destructive course to a planet that is intelligently managed" — the intelligence reflecting not only what we have between our ears, but the intelligence of super-sophisticated computers.

To illustrate his theme, Martin deals with familiar topics such as population growth, economic advance, poverty, globalism, energy, endocrine disruptors, pandemic diseases, eco-affluence, anarchic violence, terrorism and environmental ruin. He also considers issues that have received less attention, such as hydroponics, pebble-bed nuclear reactors and electronic extensions to brains, which could lead to a huge leap in human evolution.

He analyses research breakthroughs such as biotechnology and nanotechnology, along with newcomers such as regenerative medicine, extreme-bandwidth networks, and intense forms of computerized intelligence. He asserts that "computers will become 'intelligent' in a manner quite different from human intelligence. That intelligence will feed on itself, becoming more intelligent at a rapidly accelerating rate, until there is a chain reaction of computer intelligence, which we term the Singularity" — a theme that is explored in Ray



A growing problem: the economic rise of China is likely to increase its prodigious appetite for fossil fuels.

Kurzweil's recent book, *The Singularity is Near* (see *Nature* 440, 421; 2006). Martin is so optimistic about this that he thinks future research developments could prove to be far more significant than those of the Renaissance or the Enlightenment.

In order to attain a comprehensive understanding of the challenges ahead of us, Martin has donated US\$100 million to the University of Oxford, UK, to fund a new school with goals and agendas that are effectively set out in this book. Its overriding purpose will be "to identify and find solutions to the biggest challenges facing humanity in the 21st century, and to find the biggest opportunities".

The book treats many of the topics listed above, plus dozens of others, in sufficient detail to be comprehensible to the non-expert reader. All are presented as key components of the "high-culture civilizations" and "magnificent lifestyles" that Martin foresees for this century — provided humankind chooses to stay around for long enough. He agrees with Martin Rees, president of the Royal Society in London, that humankind probably has only a 50/50 chance of surviving the century: we have the capacity for unimaginable human advancement, but if we fail to address the key issues we might face disruption that could set us back by centuries — or even eliminate

us altogether. As Martin puts it, the present century is "crunch time".

Early in the book, Martin identifies a number of major trends that have unstoppable momentum, like a freight train but on a giant scale. Among such trends are population growth, tropical deforestation, fisheries decline, mass extinction of species, climate change, and others that will be readily recognizable to readers of *Nature*. Plenty of experts know how to make the train go faster, but not enough are concerned with where it is heading, or whether we want to go there anyway. Occasionally a momentum trend is brought to a sudden halt, and the result seems to come as a surprise. For decades we poured pollutants into the atmosphere, and as long as biotas seemed to have no trouble absorbing them, no one bothered to ask whether there could be any long-term repercussions. Eventually the buffering capacity of biotas suffered a critical loading of pollutants, and we professed to be surprised by the resulting acid rain. Many such surprise events, known as discontinuities, surely lie ahead, and our scientific skills in anticipating them are meagre at best, with little research being directed at these 'unknown unknowns'. Could it be that we are programmed for denial?

Some of Martin's proposals are controversial and are likely to provoke intense debate. For

LIU JIN/AFP/GETTY IMAGES

instance, he points out that many economists predict that US productivity could enjoy a long-term increase of some 2.5% per year. If extended for 100 years, this would make Americans 12 times richer in real terms than they are now. China, starting from a lower base and with a rather higher growth rate for a few decades, would be 20 times richer. Both these countries have prodigious appetites for more resources, notably fossil fuels. The world is not running out of fossil fuels yet, but it is running out of the atmospheric capacity to absorb pollutants. Will there be a radical shift in production patterns, triggering a parallel shift in consumption patterns — or the other way around? Will people have to countenance a change from 'more is better' to 'enough is best'?

A seismic transition of that order would raise all manner of questions about society and its values. "The forces of the near future are so large that they will inevitably change civilization. Human survivability (more cogent than sustainability) and creating new concepts of civilization are inextricably linked," Martin asserts.

Apocalyptic as some of Martin's writing is,

he preaches not so much 'doom and gloom' as 'doom or boom'. He sees humans "unlocking formidable new capabilities that could lead to more exciting lives and glorious civilizations" with "higher levels of happiness". Such rhetorical flourishes — there is purple prose at many places in the book — may not be to the taste of all readers, but the writing style is justified, Martin believes, by the nature of his message. He predicts an apogee of human creativity by the time of the grand transition, around the middle of the century, when humans could be standing on intellectual tiptoe to an extent never presaged until now. Indeed, it takes someone of Martin's vision to spell out such an exuberant prospect — hence the colourful phrasing to match the message. So persuasive is Martin that one can readily agree with him that, in the light of the sheer intensity of scientific research today, and of our apparent newfound capacity to solve whatever problems afflict us, the twenty-first century must surely rank as by far the finest time to be alive. ■

Norman Myers is an honorary visiting fellow at Green College, Oxford University, Upper Meadow, Quarry Road, Oxford OX3 8FS, UK.

this introduction, by reviewing, tagging and otherwise pointing the way to previously unheralded content.

Some of this content will be mainstream works that have fallen out of the public's gaze, or that may be better known in one corner of the world than another, perhaps leading a new reader to gleefully consume Jerome K. Jerome's 1889 bestseller *Three Men in a Boat*. But using the example of the amateur astronomers who observed Supernova 1987A, Anderson also argues that the ready availability of inexpensive technology is enabling many individuals to participate in areas that were previously restricted to professionals. Internet blogs, digital photos, audio programmes and video journals all add to the new marketplace of ideas.

The author has done substantial research and analysis, and refers to some earlier thinkers. However, in the chapter on the 'new producers', I was surprised that there was no reference to Alvin Toffler, whose concepts of the rise of amateur production and consumers-as-producers (which he termed 'prosumers') were presented several decades ago, notably in *The Third Wave* (Bantam, 1980). Also, an examination of the role of the public in the creation of the *Oxford English Dictionary* could have usefully contextualized the discussion of the 'Wikipedia phenomenon', in which many amateur contributors collaborate to produce encyclopaedic articles.

The concept of the long tail can be extended beyond the realm of entertainment. Anderson does this briefly, but still remains primarily in the Internet and software domains, looking at its applicability to the business of eBay and Google, among others.

I found it striking that this explosion in online content comes at a time when the world is experiencing substantial reductions in species diversity, and in a core aspect of culture — spoken languages. If the long tail can help to address this diversity crisis, it would certainly be invaluable for world culture. It seems to me that perhaps the Russian scientist

## The road less travelled

**The Long Tail: Why the Future of Business is Selling Less of More**  
by Chris Anderson

Hyperion: 2006. 256 pp. \$24.95

### Richard Akerman

The Internet has changed from a communications tool available to a small number of academics to a worldwide medium for commerce and information. Initially, business learned how to use it from the academics, but perhaps now the research community can benefit from some of the consumer-driven developments.

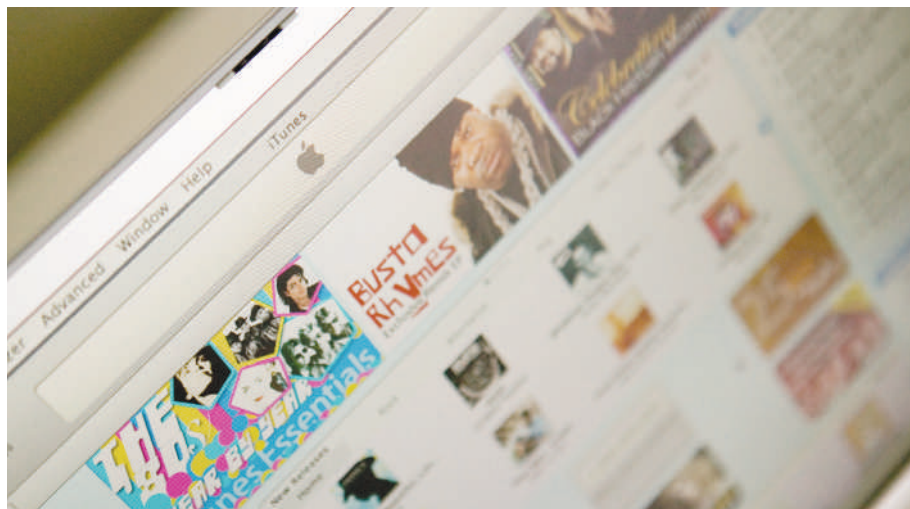
Amazon and other online retailers have discovered that there is a substantial amount of money to be made from selling a wide range of books, songs and other items that would be unprofitable or impractical to provide in physical stores, which tend to focus instead on the bestsellers. This group of less-popular content was first described by Chris Anderson as 'the long tail' in an article for *Wired* magazine in October 2004.

In his book also called *The Long Tail*, Anderson further explores the business and cultural opportunities emerging thanks to the availability and discoverability of huge amounts of entertainment and information online. He writes for and about business, and his main examples are drawn from a small number of major US Internet-centric content-distribution services including Amazon, the Rhapsody music distribution service, iTunes and the DVD rental service Netflix.

But the book also examines related Internet

developments in order to understand the multiple factors working to make the 'long tail' possible. In the electronic world, availability may be easiest of all — the cost of providing a new music track is essentially zero if a company is already storing millions of tracks. However, having provided new offerings, the company requires an audience that has somehow discovered content far beyond the heavily promoted hits they are most familiar with.

Anderson refers to this as "connecting supply and demand, introducing consumers to these new and newly available goods and driving demand down the Tail". He suggests that other consumers may be a major force in



Off the beaten track: online stores cash in by selling less-popular items as well as those by big names.



Nikolai Ivanovich Vavilov, travelling the world in the early twentieth century to amass a vast collection of plant seeds, was an early explorer of the long tail of plants.

From the perspective of scholarly communication, librarians, publishers and academic institutions have much to gain from the ideas

in this book. Scientific content, such as data sets and peer-reviewed papers, is being produced at a breathtaking pace, yet there are still gaps in terms of availability and discoverability. The techniques used to locate previously hard-to-obtain music tracks, obscure books and unexpected content relationships can be

adapted to enable researchers to discover less-cited but potentially useful data and papers. ■ Richard Akerman is at the Canada Institute for Scientific and Technical Information, Ottawa, Ontario K1A 0R6, Canada.

**Supplementary links for this review are available at [www.connotea.org/user/scilib/tag/longtailreview](http://www.connotea.org/user/scilib/tag/longtailreview)**

## Sensitive to modern life

### **Allergy: The History of a Modern Malady**

by Mark Jackson

Reaktion: 2006. 256 pp. £25, \$39.95

#### **Peter J. Barnes**

The term 'allergy' was first used by the Austrian paediatrician Clemens von Pirquet in 1906. He was describing an exaggerated biological reactivity to foreign substances, which he demonstrated by means of injections. Asthma, eczema and hay fever had been recognized since antiquity, but in the early 1900s these allergic diseases were considered uncommon. Their increase in prevalence in the past century is unlikely to be due to better diagnosis, as they are usually easy to recognize. In the past 20 years in particular there has been a dramatic increase in allergic diseases of all types in industrialized countries. This is much too rapid to be accounted for by genetic changes, although we know that genetic predisposition is important in the development of allergies, so there must be an environmental cause. The increase in allergic diseases seems to relate most closely to the adoption of a Western lifestyle.

In *Allergy*, Mark Jackson, professor of the history of medicine at the University of Exeter, UK, traces the historical development of allergy, discusses possible explanations for the recent explosive increase, and concludes that it is a disease of civilization. The book provides a perceptive insight into the historical development of allergy, indicating how thinking changes. It gives fascinating vignettes of key researchers involved in the history of allergy and contains some interesting anecdotes about their lives.

Various environmental causes have been suggested for the increase in allergic diseases, including changes in diet, poorly ventilated housing, increased exposure to allergens, air pollution (commonly believed by patients to be important) and, most convincingly, reduced infections during early childhood, which favours an allergic pattern of immunity. No single factor accounts for the rise, however, which is likely to be caused by a combination of environmental factors that occur together in Western society. Asthma, allergic rhinitis (including hay fever) and eczema have all increased, suggesting that they have an underlying allergic mechanism. Many epidemiological studies have now been done, but it is still not known exactly which environmental



**Masks and goggles are the order of the day as Japan struggles to cope with a rise in hay fever.**

factors are important in the development of allergies in genetically predisposed patients.

The most compelling evidence for an environmental cause is the high prevalence of allergies in the former West Germany compared with a very low prevalence in the East. These differences are now disappearing in the united Germany as the environmental conditions become more similar. Other convincing evidence is the low prevalence of allergic diseases in children brought up on animal farms, where there is a high level of exposure to endotoxins, which stimulate a protective pattern of immunity. The lack of infections and endotoxin exposure associated with improved hygiene and the widespread use of antibiotics fails to stimulate this protective immunity. Jackson discusses these theories, although he does not provide much evidence for and against each idea.

Allergies are costly, in terms of medication, hospital admissions from asthma, and time taken off work. It is estimated that allergic diseases cost more than £1 billion (US\$1.9 billion) in Britain alone, excluding hospital costs. The good news is that the rise in allergic diseases seem to be slowing down. In some countries, 40–50% of allergy skin tests are positive, and this may be as high as the genetic predisposition will allow. Allergic diseases have been very profitable for pharmaceutical companies, and asthma medications are one of the fastest-growing markets in the world. But the high costs of medication have promoted a search for non-pharmacological approaches centred on changing the environment. Allergen

avoidance has usually been disappointing, as it has been impossible to completely avoid exposure, and changes in diet (more antioxidants and increased consumption of oily fish) have not been effective.

Perhaps the most promising approach to allergic disease is vaccination, which is particularly effective against hay fever and in desensitizing against insect stings. Desensitization — the administration of gradually increasing doses of an allergen to promote a long-term change in the immune response — is discussed at various points in the book. However, the treatment can sometime provoke exaggerated responses or even anaphylactic shock, prompting the search for safer approaches. Treatments placed under the tongue, rather than given as skin injections, are effective, and purified allergen proteins may be safer. In the future, DNA vaccines and T-cell peptides could provide a means of inhibiting the allergic response.

Jackson's succinct and clearly written book is aimed at the informed lay reader. He admirably avoids using jargon and scientific terminology, and gives fascinating insight into the rise in allergic diseases and how this is linked to our modern lifestyle. I recommend this book, which helps us to understand the relationship between health and the environment, and explains why modern living can be detrimental to our health. ■

Peter J. Barnes is head of respiratory medicine, National Heart and Lung Institute, Imperial College London, London SW3 6LY, UK.

# Hearing colours, seeing sounds

Wassily Kandinsky's synaesthetic paintings go on show in London.

**Martin Kemp and Colin Blakemore**

"I saw all my colours in spirit, before my eyes. Wild, almost crazy lines were sketched in front of me."

The author was not describing effects he had seen with his eyes. Wassily Kandinsky was reacting to a performance of Richard Wagner's opera *Lohengrin* in St Petersburg. The Russian painter and art theorist, who spent much of his career in Germany, conceived a kind of painting that might aspire to the abstract condition of music. His aim was nothing less than a new 'science of painting' that would exploit the inherent power of form and colour.

'Colour music' has a long tradition. Aristotle and his followers in ancient Greece provided encouragement for artists and theorists from the Renaissance onwards to explore precise connections between 'colour scales' (arrayed between the poles of white and black) and the mathematical harmonies of music. Attempts to devise detailed colour notations for music date back to the sixteenth century, and the earliest attempt to construct a 'colour organ' was made as long ago as the early eighteenth century.

The early twentieth century saw a cluster of diverse attempts to weld the visual and aural into an abstract unity. Kandinsky moved freely in the intellectual circles of two composers, Alexander Scriabin and Arnold Schoenberg, who were openly involved in this endeavour. The score of Scriabin's *Prometheus: Poem of Fire* (1910) includes a part for a 'luxe' or 'clavier à lumières', which was intended to project a 'symphony' of vivid colours on a screen.

Kandinsky's *Über das Geistige in der Kunst* (Concerning the Spiritual in Art, 1911) draws on the ideas of theosophy, the philosophy that advocates a form of pantheistic spiritual insight based on the deepest affinities between world theologies and the principles of science. He advocated a "renewal of the soul" in the face of rampant materialism.

His personal contribution was an art that relied on a direct communication of 'inner vibrations', in which visual arrays of shapes and colours evoke sensory conjunctions — embracing not only sound but also taste, smell and touch. Colour serves as the

"keyboard, the eyes are the hammers, the soul is the piano with many strings". Looking at photographs of *Composition VII* during progressive stages of execution, Kandinsky's gestures with loaded brush in hand look like those of a conductor drawing phrases and colours from an orchestra — although his visual sounds endure, not fade away.

Kandinsky realized that all effects are relative. Each rhythmic stroke, each adjacent colour, is part of a dynamic system of pulsing interaction. Signature circular patches radiate haloes of applied and induced

synaesthetes listening to words that evoked synaesthetic colour sensations.

Colour is the 'added value' of vision. Numerous studies of human vision suggest that colour is separately analysed and then attached to the perception of objects and surfaces. Perhaps the prevalence of colour metaphors, and the fact that colour is so frequently the extra experience of synaesthetes, reflect a normal mechanism for associating colours with objects.

Kandinsky's writings suggest that he did experience coloured music. His distinctive,

brightly coloured swirling and jagged lines might be representations of experiences evoked by music. But his approach may have been more formal. Kandinsky left an annotated copy of Schopenhauer's treatise on vision and colour, *Über das Sehn und die Farben* [*On Vision and Colour*], with his lover, artist Gabrielle Münter, when he left her in 1914. This was the year he painted *Fugue*, whose interwoven sequences of coloured patterns were explicit representations of musical motifs.

Beyond such speculation about Kandinsky lies the intriguing fact that colours have pervasive emotional associations for all of us. Although explicit synaesthesia is uncommon (affecting about 1 in 2,000 of the population), about one-third of non-synaesthetic people are roughly 70% consistent in associating colours with graphemes in separate tests a week or a month apart (M. S. Steven D. Phil. thesis, Univ. Oxford, 2004).

Regardless of where Kandinsky stood in the spectrum of synaesthetes, he has tapped into one of the stranger capacities of our perceptual system, creating works that were in their own way as experimental as those of a scientist. The key difference is that his experiments are presented as open fields for our imaginative viewing, rather than as parades of results of controlled testing.

An exhibition of his paintings, 'Kandinsky: The Path to Abstraction', can be seen at Tate Modern in London until 1 October, and at the Kunstmuseum in Basel from 21 October to 4 February 2007.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, and Colin Blakemore is Waynflete professor of physiology at the University of Oxford, Oxford OX1 3PT, UK.



colours. Shallow curves arch harmonically, only to be interrupted by jagged dissonances. *Composition VII* (pictured here), 2 × 3 metres in size, invites us to read its 'music' over time in an act of sequential contemplation.

This cross-sensory endeavour raises the question of whether Kandinsky experienced synaesthesia, in which particular sensory stimuli trigger extraneous sensations. As with Scriabin, the answer is not straightforward. Synaesthesia is thought to be disproportionately common among creative artists, but they were both knowing heirs to the long tradition of colour music (see chapter 11 of John Gage's *Colour and Culture*, Thames & Hudson, 1995).

In synaesthesia, all manner of weird associations can occur (such as imaginary tastes associated with the sight of particular people), but the commonest is 'colour grapheme' synaesthesia, in which written or printed letters and numbers appear distinctively coloured. We don't come into the world knowing about graphemes, so the particular associations must be learned, even though synaesthesia itself is partly heritable. Using functional brain imaging, the colour-processing areas of the visual cortex have been seen to light up in



## EVOLUTIONARY BIOLOGY

# How to build a longer beak

Nipam H. Patel

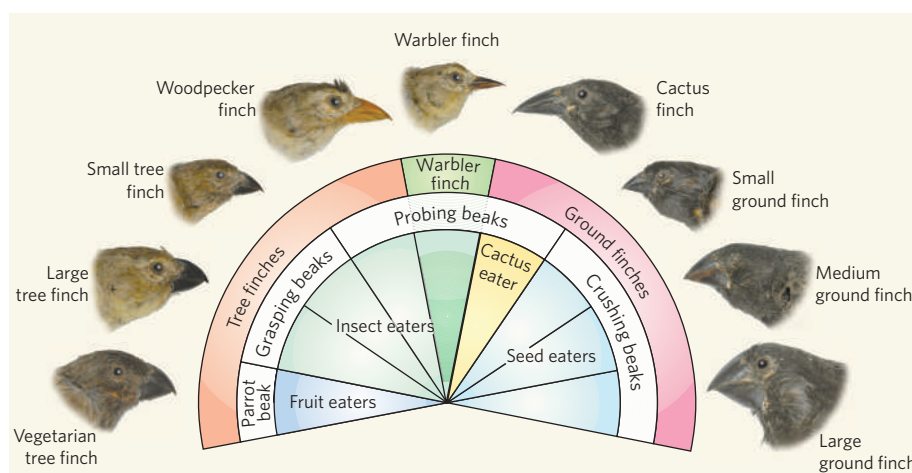
**Evolutionary changes in the beaks of Darwin's finches have been instrumental in the adaptive radiation of these birds. The molecular basis for variation in beak size and shape is opening up to investigation.**

A classic illustration of nature's ability to generate morphological diversity comes from the finches that inhabit the Galapagos Islands. The beak shapes of these finches are remarkably diverse, and — as described on page 563 of this issue<sup>1</sup> — Abzhanov and colleagues have uncovered one of the mechanisms involved in achieving this. They have compared beak development of several finch species at the molecular level. In combination with experimental analyses in chickens, they show that changes in calcium-dependent molecular signalling during development are involved in the evolution of beak shape.

When Darwin visited the Galapagos Islands, he collected several birds that are now known as Darwin's finches. There are roughly a dozen species of them, and we recognize today that they are closely related to one another, despite their remarkable differences in beak shape and size, and in eating habits (Fig. 1). Indeed, they are so varied that Darwin did not immediately realize that they were all finches, and did not initially use them to support his emerging theories of evolution.

Subsequently, however, the finches have figured prominently in our understanding of the mechanisms of evolution. In particular, over the past 30 years or more, Peter and Rosemary Grant have shown that variation in the finches is driven by ecological forces, and that the birds' adaptive radiation — their rapid speciation from a common ancestor to fill many ecological niches — has occurred in just the past few million years. The precise dimensions (length, depth and width) of each species' beak are crucial to their lifestyle and survival<sup>2</sup> (Fig. 1), and fluctuations in the environment lead to selection that changes the relative success of birds with various beak shapes. Indeed, these evolutionary processes are evident in real time on the Galapagos Islands<sup>3</sup>.

An initial insight into the molecular and genetic underpinning of the different beak morphologies came from Abzhanov and colleagues two years ago<sup>4</sup>. They focused on genes that regulate the skeletal and cartilaginous development of the face in laboratory species such as mice and chickens. Using this 'candidate gene approach', they showed that a gene



**Figure 1 | The evolutionary radiation of Darwin's finches.** The species depicted here have all arisen from a common ancestor, but have evolved a remarkable diversity of beak shapes and sizes as they have adapted to exploit different food sources. The work of Abzhanov *et al.*<sup>1</sup> implicates upregulation of the calmodulin-dependent signalling pathway in the evolution of the long, pointed beak of the cactus finch. (Finch specimens provided by the Museum of Vertebrate Zoology, Berkeley.)

encoding a cell–cell signalling protein, known as bone morphogenetic protein 4 (BMP4), was more broadly expressed during the embryonic development of the deep and wide beaks of ground finches than during the development of finches with narrower beaks. To go beyond correlation, they then showed that broadening the expression of the gene encoding BMP4 in chicken embryos resulted in chickens with wider beaks.

The candidate-gene approach, however, is restricted to the relatively few genes known to play a role in facial development. In their latest research<sup>1</sup>, Abzhanov *et al.* used a microarray approach to look, in a relatively unbiased manner, for genes that are differentially expressed between species and that are correlated with beak morphology. In this way, they identified a gene that encodes a protein, called calmodulin (CaM), that is expressed at much higher levels in the long and pointed beaks of cactus finch embryos than in the beaks of other finch embryos. CaM mediates calcium signalling in cells, which in turn plays many roles in controlling cell and tissue differentiation and development. Thus, this part of the analysis provided a striking correlation — high levels

of CaM are associated with the long and pointed beaks of cactus finches.

To go beyond correlation, however, Abzhanov *et al.*<sup>1</sup> used the biochemical knowledge of the CaM-dependent signal-transduction pathway. In response to elevated levels of calcium in the cell, CaM is activated, and in turn activates a protein known as CaMKII, which then activates many other known target proteins. Previous studies had identified a mutant form of CaMKII that is constitutively (continuously) active, and by expressing this mutant form in the beak of chicken embryos, Abzhanov and colleagues believed they could generate the same sort of change in signalling that results from raising levels of CaM expression.

When they elevated CaM-dependent signalling in this way, they created chickens with longer, but not wider or deeper, beaks. They were thus able to create a change that reproduced at least part of the evolutionary distinction seen between cactus finches and other finches. Combined with the work on BMP4, these experiments also show that different aspects of beak size (length versus width and depth) are controlled separately at the genetic



## 50 YEARS AGO

"Nuclear magnetic resonance and electron spin resonance" —

It is no new idea that most of the advances in the techniques of chemistry have come from the use of apparatus and methods originally devised by physicists. This is particularly true of spectroscopy, where the main applications associated with the different frequency bands have moved over, one after another, from the pure physicist to the chemist... It would appear that during the past eighteen months a similar move has been taking place with the two new techniques of nuclear magnetic resonance and electron spin resonance... As with all techniques which study the interaction of atoms with external forces, it soon became clear that these new methods also had very great potentialities as tools for chemical investigation, and during the past few years these applications have been brought to light in a very striking way. **D. J. E. Ingram**

From *Nature* 4 August 1956.

## 100 YEARS AGO

"Strength of a Beetle" — Last night a small beetle (*Aphodius fossor*), the length of which is  $\frac{1}{2}$  inch, flew in at my window and alighted on a table next to me. As it buzzed about I put a lid of a tin box over it, but to my surprise the beetle walked about bearing the lid on its back. I then put the tin box on top of the lid, and was absolutely amazed to find that the insect tilted up a corner of the combined box and lid, and nearly escaped. The weight of the beetle when dead was  $\frac{1}{2}$  grain, alive I suppose it was a little more; but the box and lid weighed 1758 grains! Assuming that the living insect weighed 1 grain, it must have tilted up 1758 times its own weight! Of course, the strength required to tilt up a box on edge is nothing like so great as that required to actually lift the weight, but nevertheless the feat seems to me sufficiently astounding. The dimensions of the box are  $3\frac{1}{8} \times 2\frac{1}{8} \times 1\frac{1}{2}$  inches. From *Nature* 2 August 1906.

level, allowing each dimension to evolve independently.

However, these results do not establish the precise nature of the genetic changes responsible for the differences in morphology. Are the changes in CaM levels between species due to one or more differences in the CaM gene itself — for example in the flanking regulatory regions of the genome that control where and at what rate the gene is transcribed? Or are they due to changes in one or possibly many genes scattered throughout the genome that act upstream of CaM to cause it to be expressed at higher levels in the developing beaks of cactus finches?

Genetic mapping studies in other animals and plants, such as maize (corn) and teosinte (from which maize was domesticated)<sup>5</sup>, suggest that mutations directly affecting the expression level of a single gene have been responsible for some profound evolutionary changes. The application of mapping techniques could

answer this question in Darwin's finches. Such information would add to the debate over whether evolution proceeds through the accumulation of many mutations of small effect in many genes, or through one or a few mutations of large effect in a single gene. Although it is difficult to generalize from a few examples, Darwin's finches still have much to tell us about the evolutionary process. ■

Nipam H. Patel is in the Departments of Molecular and Cell Biology, and Integrative Biology, and at the Howard Hughes Medical Institute, University of California, Berkeley, California 94720-3140, USA.  
e-mail: nipam@uclink.berkeley.edu

1. Abzhanov, A. et al. *Nature* **442**, 563–567 (2006).
2. Grant, P. R. *The Ecology and Evolution of Darwin's Finches* (Princeton Univ. Press, 1999).
3. Grant, P. R. & Grant, B. R. *Science* **313**, 224–226 (2006).
4. Abzhanov, A. et al. *Science* **305**, 1462–1465 (2004).
5. Doebley, J., Stec, A. & Hubbard, L. *Nature* **386**, 443–445 (1997).

## PARTICLE PHYSICS

# A finer constant

Andrzej Czarnecki

**For the first time in a decade, the precision of the fine-structure constant — central to understanding the electromagnetic force — has improved. But even greater accuracy is required to test new physics.**

How does the colour of a rose relate to the hardness of oak? To physicists, both result from electromagnetism, an interaction whose strength is encoded in one pure number — the fine-structure constant. Appropriately for something introduced at the dawn of quantum mechanics, the fine-structure constant is denoted by the Greek letter alpha ( $\alpha$ ). It was once believed to be a simple fraction,  $1/137$ , a circumstance that provoked theorists to search for some deeper meaning to it. Studied closer, the denominator turned out not to be an integer. Writing in *Physical Review Letters*, Gabrielse and colleagues<sup>1</sup> use a measurement of the electron's magnetic moment reported in a companion paper<sup>2</sup> to find that  $\alpha = 1/137.03599710(96)$ , the most accurate value yet. But why is this important — and why is even this accuracy not enough?

Electromagnetism dominates most phenomena at scales larger than the subatomic (which is ruled by nuclear forces) but shorter than the astronomical (the realm of gravity). Thus,  $\alpha$  can be measured in many ways, using any system of well-understood electromagnetic nature. When Arnold Sommerfeld first used  $\alpha$  in 1915, he named it the fine-structure constant because it described subtle features of the radiation spectrum of the hydrogen atom. Its value was initially best determined by measuring atomic transitions. In the 1970s, more

precise values came from solid-state systems, through the discovery of electrical phenomena such as the Josephson and the quantum Hall effects (Fig. 1).

For the past quarter-century, the world record for the most accurate value of  $\alpha$  has been held by amazing experiments performed on a single electron trapped in a vacuum permeated by electric and magnetic fields<sup>3</sup>. The electron, as a charged and rotating particle, is a tiny magnet with a strength — its magnetic moment — given by  $\mu = g(e/2m)s$ , where  $e$ ,  $s$  and  $m$  are the electron's charge, spin and mass. The proportionality coefficient  $g$  would be 1 for a classical spinning ball. For the point-like electron, relativity theory demands that  $g = 2$ .

This is not yet the whole story. The physical vacuum, far from being 'nothing', vibrates with activity. Elementary particles borrow energy from the vacuum to pop up and disappear again through quantum fluctuations. The electron interacts with such 'virtual' particles, mainly photons, and its  $g$ -factor is increased slightly by an amount that depends on  $\alpha$ . This deviation, known as 'g minus two' ( $g - 2$ ) is among the most precisely calculated quantities in physics. In fact, quantum electrodynamics, the theory of electron interactions with light, was born through efforts to understand its value.

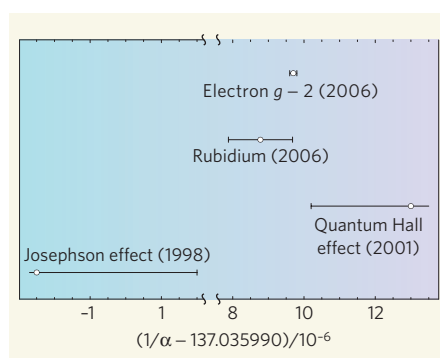
The dream of the theorist is an exact expression for  $g - 2$  in terms of  $\alpha$ , but that seems as



elusive as deriving  $\alpha$  itself. For now, only the first four terms of the Taylor expansion of  $g - 2$  are known (this is a mathematical expression consisting of a series of terms that, added together, come closer and closer to the true value of a quantity). Getting that far took six decades and spawned a rich tool-box of mathematical methods and tricks that has benefited other branches of science. Not least among these is the field of symbolic computation, which aims to harness the power and patience of computers to tackle huge algebraic equations.

This theoretical progress has gone hand in hand with experimental breakthroughs. Whereas particle physicists have consistently strived to build higher-energy accelerators, Gabrielse and Peil at Harvard University succeeded in constructing a cyclotron with the lowest energy so far<sup>4</sup>. In this cavity, cooled to 100 millikelvin, a single electron can be trapped and screened even from some of the vacuum fluctuations. Such is the tranquillity of its setting that individual quantum levels of the electron motion and spin can be discerned. Transitions among the lowest-lying states have now allowed Gabrielse and colleagues to determine the  $g$ -factor to an accuracy of one part in a trillion (ref. 2) and, when compared with the theoretical expression, improve our knowledge of  $\alpha$  (ref. 1).

Can the accuracy of  $\alpha$  be improved indefinitely? This is unlikely using the electron  $g - 2$ . Gabrielse and colleagues' measurement<sup>1</sup> is, for the first time, sensitive not only to electromagnetic forces but also to tiny, strong nuclear effects. This is an impressive achievement, but also hints at obstacles ahead: the uncertainty inherent in nuclear forces will



**Figure 1 | Approaching alpha.** The fine-structure constant,  $\alpha$ , is present wherever there is electromagnetism. The accuracy of its determination reflects our degree of understanding of various phenomena. Gabrielse and colleagues' value<sup>1</sup> using the electron  $g$ -factor is the best yet; other sources include photon collisions with rubidium atoms, the Josephson effect and the quantum Hall effect. Neutrons, caesium and several other atomic systems have also been used.

eventually limit how well  $\alpha$  can be read off from  $g - 2$ .

But this emerging sensitivity to physics beyond electromagnetism is creating new opportunities. The electron has a heavier cousin, the muon, whose  $g - 2$  has recently been measured. The result disagrees<sup>5</sup> with our understanding of fundamental interactions by a tantalizing 2.8 standard deviations: too much to ignore, but not enough to claim a discovery of something new wiggling in the vacuum. Nevertheless, this could be the first evidence of particles that are too massive to have been seen in our laboratories, but that were perhaps crucial in the design of the Universe.

Could measurements of the electron help clarify what has or hasn't been seen? To probe the putative new force the muon might have sensed, the accuracy of the electron  $g - 2$  measurements should be improved by a factor of at least a dozen. Of course, an independent value for  $\alpha$  would be needed to interpret the result.

Excitingly, atomic studies — the original source of information on  $\alpha$  — are currently seeing a renaissance, and could produce sufficiently precise values. The new laser-based tools of optical lattices and frequency combs, recognized with the 2005 Nobel Prize in Physics, have been applied<sup>6</sup> to trap rubidium atoms and thus determine  $\alpha$  with an error of seven parts in a billion. This is about nine times cruder than Gabrielse and colleagues' value<sup>1</sup>. But if the expected improvements in the atomic approach materialize, the electron  $g - 2$  will be freed to check out new physics.

Improving our knowledge of the fine-structure constant by another order of magnitude with an independent method is a daunting task. But the cunning of the  $\alpha$  hunters justifies cautious optimism. ■

Andrzej Czarnecki is in the Department of Physics, University of Alberta, Edmonton, Alberta T6G 2J1, Canada, and the Tri-University Meson Facility (TRIUMF), Vancouver, Canada. e-mail: czar@phys.ualberta.ca

1. Gabrielse, G., Hanneke, D., Kinoshita, T., Nio, M. & Odom, B. *Phys. Rev. Lett.* **97**, 030802 (2006).
2. Odom, B., Hanneke, D., D'Urso, B. & Gabrielse, G. *Phys. Rev. Lett.* **97**, 030801 (2006).
3. Schwinger, P. B., Van Dyck, R. S. Jr & Dehmelt, H. G. *Phys. Rev. Lett.* **47**, 1679–1682 (1981).
4. Peil, S. & Gabrielse, G. *Phys. Rev. Lett.* **83**, 1287–1290 (1999).
5. Bennett, G. W. *et al.* *Phys. Rev. Lett.* **92**, 161802 (2004).
6. Cladé, P. *et al.* *Phys. Rev. Lett.* **96**, 033001 (2006).

## CHEMICAL BIOLOGY

# Cutting out the middle man

Tom W. Muir

**Wouldn't it be nice if you could control the function of any protein with one small molecule? Unlikely as it sounds, this could become possible through a crafty process known as protein splicing.**

One of the most exciting things about chemical biology is its potential to develop new tools for probing cellular processes. Small 'drug-like' molecules that can diffuse into cells and quickly elicit a discernible response offer distinct advantages over genetics-based approaches for exploring the highly choreographed inner workings of a cell<sup>1</sup>. In particular, small molecules can allow cellular processes to be rapidly perturbed in a reversible and often tunable fashion, allowing the dynamic features to be teased out. But finding a small-molecule modulator that is specific for one protein is a formidable challenge; finding one molecule

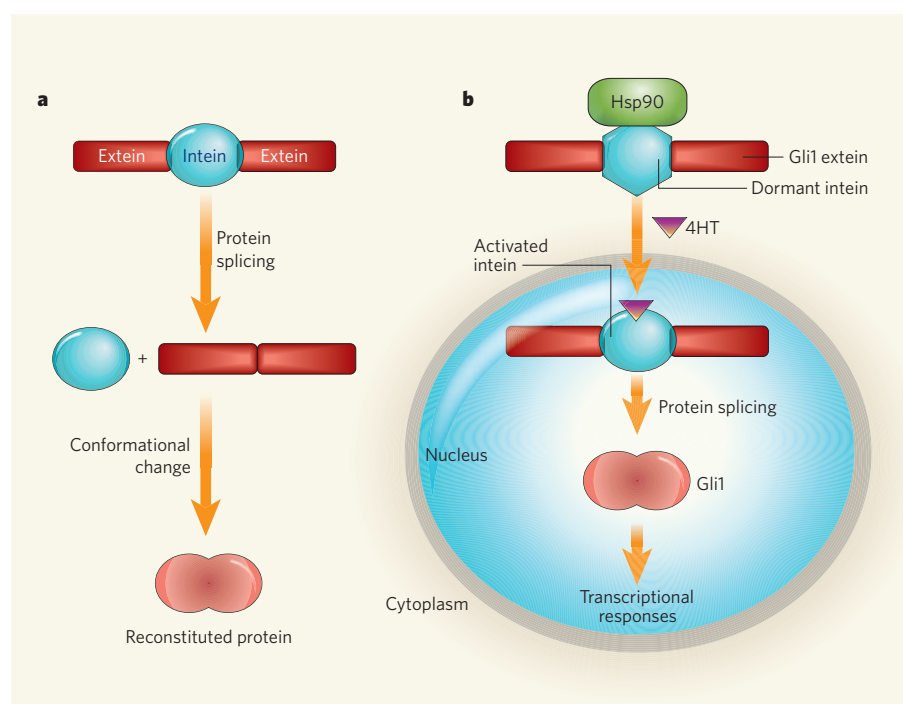
that could somehow modulate the function of any desired protein seems impossible. Yet in the *Journal of the American Chemical Society*<sup>2</sup>, Yuen *et al.* describe the results of a study, based on the pharmacological regulation of protein splicing, that suggest this might not be a complete pipedream.

Protein splicing is one of the most dramatic protein modifications known<sup>3</sup>. It is a self-catalysed process in which an internal protein domain, known as an intein, removes itself from a host protein with concomitant linking together of the flanking polypeptides, the exteins (Fig. 1, overleaf). Some inteins are

remarkably promiscuous with respect to the extein sequences within which they sit. Indeed, inteins have found widespread applications in protein engineering as a way to introduce biochemical and biophysical probes into proteins<sup>4</sup>.

Protein splicing results in a new polypeptide sequence being produced at the site of intein excision. Because a protein's function is intimately linked to its sequence, splicing has the potential to regulate the activity of the host protein. With this in mind, conditional inteins have been reported whose splicing activity is triggered by changes in temperature<sup>5</sup> or by the application of small molecules<sup>6–8</sup>. The appeal of these systems is that inducible inteins can be dropped into target proteins with standard molecular-biology techniques.

The work of Yuen *et al.*<sup>2</sup> builds on a previous study<sup>7</sup> from the same group, in which directed protein evolution was used to develop a controllable protein-splicing element that could be activated by the addition of a ligand — a molecule that binds to the protein. The controllable intein was based on a hybrid

**Figure 1 | Mechanism of protein splicing.**

**a**, In protein splicing, an intein spontaneously removes itself from a host protein. The flanking polypeptides, known as exteins, are concomitantly linked together, leading to reconstitution of protein function. **b**, Small-molecule-controlled protein splicing uses an intein (blue hexagon) that is activated by 4-hydroxytamoxifen (4HT). The intein-containing protein (in this case, Gli1) is initially associated with the cytoplasmic protein Hsp90. Addition of 4HT releases the Gli1-intein protein from Hsp90, and initiates protein splicing, most likely as a result of a conformational change of the Gli1-intein protein (the hexagon turns into an oval). Protein splicing occurs in either the cytoplasm or the nucleus. Gli1 is produced, and regulates the transcription of DNA associated with embryonic development<sup>2</sup>.

formed between the ligand-binding domain of the oestrogen receptor (ER-LBD) and the *Mycobacterium tuberculosis* RecA intein. Addition of the high-affinity ER-LBD ligand 4-hydroxytamoxifen (4HT) led to dose-dependent, rapid splicing of the engineered intein out of a series of model proteins. This work<sup>7</sup> was performed in yeast, so it was unclear whether the technology would work in a mammalian cell, or whether it could be used to control an actual biological process. These questions are addressed in the new study<sup>2</sup>.

The authors began by proving that they could induce splicing in mammalian cells, and not just in yeast. They inserted their controllable intein into a jellyfish protein, known as green fluorescent protein (GFP) because it fluoresces green when exposed to blue light. The intein-containing protein was expressed in mammalian cells, and was treated with 4HT to induce splicing. As had been hoped, the splicing product was found to have native GFP properties, as indicated by fluorescence. Splicing was not observed in the absence of 4HT and the amount of fluorescence correlated well with the dose of 4HT administered. Interestingly, the GFP splicing reaction was considerably slower than the same process in yeast. This may be because growth conditions and cell physiology differ in yeast and mammalian cells. Further optimization of the system may be possible through additional rounds of protein evolution<sup>2</sup>.

Next, the authors asked whether the splicing technology can be applied to native mammalian proteins. They inserted the controllable intein into the DNA-binding domains of Gli1 and Gli3, important mediators in mammalian embryo development. These proteins regulate transcription — the first stage of gene

expression, in which genetic information is transferred from DNA to RNA. Gli1 activates transcription, whereas Gli3 inhibits it. Splicing of both these intein-containing proteins was found to be dependent on the addition of 4HT. Moreover, the proteins move from the cell cytoplasm to the nucleus in response to 4HT. Studies indicate that addition of 4HT leads to the expected activation of transcription for the Gli1-intein and repression of transcription for the Gli3-intein.

ER-LBD is known to interact with the cytoplasmic protein Hsp90, an association that is disrupted by 4HT (Fig. 1). To rule out the possibility that the alleviation of cytoplasmic sequestration was responsible for the observed transcriptional effects, a mutant version of the intein that could no longer splice was used. Addition of 4HT to cells expressing Gli1-intein or Gli3-intein mutants did not lead to changes in transcription, although nuclear localization still occurred. This proves that protein splicing is required for the observed transcriptional responses to 4HT.

Finally, and perhaps most significantly, two versions of the controllable Gli1-intein protein were introduced into an embryonic mouse cell line. In both cases, 4HT addition led to the activation of an enzyme, alkaline phosphatase, which is a known marker of bone remodelling cells (known as osteoblasts). So, activation of Gli1-intein seems to lead to osteoblast differentiation in culture, bypassing other signalling cues that normally control this process. This striking proof-of-principle suggests that the individual processes involved in embryonic development could be activated at will, so that the resulting fate of the cells can be studied.

It would be wrong to think of controllable protein splicing as a panacea that will render

obsolete high-throughput screening initiatives for the discovery of small-molecule tools. For one thing, protein splicing is not reversible, and even in the most rapid cases<sup>7,9</sup> it is slower than traditional pharmacological manipulations. Moreover, it is unlikely that inteins will splice out of every extein context. Even when splicing can be triggered, it may not be straightforward to achieve differential activity between the pre-splicing and post-splicing states — that is, the precursor may retain some activity or the product may not regain activity. Nevertheless, a great advantage of this approach is that different constructs can be made quickly through standard molecular-biology techniques, and, as the work of Yuen *et al.*<sup>2</sup> so beautifully demonstrates, controllable protein splicing can be used to regulate the function of an essential cellular protein. As the technology continues to be refined, we can expect to see this approach used increasingly by chemical biologists.

Tom W. Muir is at the Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

e-mail: muirt@rockefeller.edu

- Shogren-Knaak, M. A., Alaimo, P. J. & Shokat, K. M. *Annu. Rev. Cell Dev. Biol.* **17**, 405–433 (2001).
- Yuen, C. M., Rodda, S. J., Vokes, S. A., McMahon, A. P. & Liu, D. R. *J. Am. Chem. Soc.* **128**, 8939–8946 (2006).
- Paulus, H. *Annu. Rev. Biochem.* **69**, 447–496 (2000).
- Muralidharan, V. & Muir, T. W. *Nature Meth.* **3**, 429–438 (2006).
- Zeidler, M. P. *et al. Nature Biotechnol.* **22**, 871–876 (2004).
- Mootz, H. D. & Muir, T. W. *J. Am. Chem. Soc.* **124**, 9044–9045 (2002).
- Buskirk, A. R., Ong, Y.-C., Gartner, Z. J. & Liu, D. R. *Proc. Natl Acad. Sci. USA* **101**, 10505–10510 (2004).
- Skratas, G. & Wood, D. W. *Protein Sci.* **14**, 523–532 (2005).
- Mootz, H. D., Blum, E. S., Tyszkiewicz, A. B. & Muir, T. W. *J. Am. Chem. Soc.* **125**, 10561–10569 (2003).



## CELL BIOLOGY

## Polarity bites

Richard Fehon

**Cells often need to have polarity to function — cells lining the gut, for instance, secrete digestive enzymes only from their intestinal side. A protein called Bitesize is pivotal in determining which way is up.**

Epithelial cells, such as those that make up the skin and the lining of the intestine, all share fundamental characteristics — the ability to stick tightly to one another and an asymmetric organization of the proteins and lipids that constitute the cell membrane (a property known as apical/basal polarity). These properties are essential for the epithelia to form protective barriers and to regulate cell proliferation and specialization. Disruption of either characteristic can have catastrophic effects — the loss of intercellular adhesion and apical/basal polarity allows cancer cells to spread, for instance. Both adhesion and polarity rely on a highly organized internal structural network called the cytoskeleton. On page 580 of this issue, Pilot *et al.*<sup>1</sup> shed light on how epithelial cells organize the cytoskeleton, and thereby regulate both assembly of specialized adhesive structures and apical/basal polarity\*.

An intricate system of interacting protein complexes divides the cell surface into at least three functionally and molecularly distinct domains: the apical membrane, the basolateral membrane and, in between, the adherens junction<sup>2</sup> (Fig. 1). The adherens junction is composed of the adhesive protein E-cadherin, which spans the cell membrane, plus associated cytoplasmic catenin proteins and actin microfilaments (which make up the cytoskeleton). The current consensus is that the different domains are characterized by the presence of particular multi-protein complexes — the Bazooka/Par-3 (Baz/Par-3) complex and the

Crumbs complex define the apical domain, and the Discs-large complex defines the basolateral domain. Genetic studies suggest that the apical and basolateral complexes compete to 'stake out' regions of the plasma membrane, with the adherens junction serving as referee by physically separating these competing parties<sup>3,4</sup>.

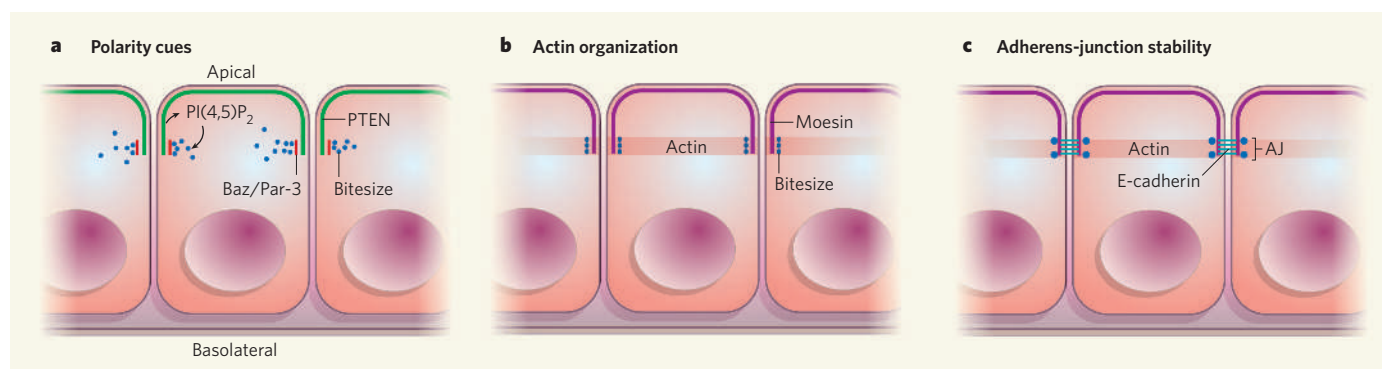
How do the complexes assemble reproducibly in specific regions of the plasma membrane? The formation of adherens junctions requires the Baz/Par-3 complex. However, this complex does not need an intact adherens junction to assemble in the apical domain of fruitfly epithelial cells<sup>5</sup>. Furthermore, intact microfilament and microtubule cytoskeletons are required for Bazooka to reach the apical membrane<sup>6</sup>. These results suggest that an inherent cytoskeletal polarity provides the initial cue for assembly of the Baz/Par-3 complex, which in turn controls the subsequent assembly of the other components that define apical/basal polarity.

Although satisfying, this model leaves several significant questions unanswered. In particular, there is the issue of how adherens-junction assembly is controlled downstream of Baz/Par-3, which Pilot *et al.*<sup>1</sup> now address. Using genome-wide arrays to pinpoint genes expressed during the development of epithelia in fruitfly embryos, they picked out *bitesize*, a gene that encodes a synaptotagmin-like protein that has previously been identified as affecting growth in adult flies<sup>7</sup>. Synaptotagmin-like proteins bind to members of the Rab family of enzymes called small GTPases

and have been implicated in exocytosis<sup>8</sup>, a process by which cells transport lipids and certain proteins to the cell surface, in many cases aided by the cytoskeleton. However, it is not known whether the Bitesize protein has a similar function.

Pilot *et al.* next examined the role of Bitesize in epithelial morphogenesis using RNA interference (RNAi). This technique targets specific messenger RNAs for degradation, thus halting production of the encoded protein. These RNAi 'knockdown' experiments, together with experiments using standard genetic mutations, showed that when *bitesize* function is disrupted, epithelial integrity is compromised. Careful microscopy of living mutant embryos established that, in the absence of Bitesize, the Baz/Par-3 complex and adherens-junction proteins are recruited normally to the apical membrane in the early stages of epithelial formation, but the adherens junction is not stable and rapidly disassembles. In addition, Bitesize requires Bazooka to localize apically, but does not need E-cadherin. These results suggest that Bitesize functions downstream of the Baz/Par-3 complex and upstream of the adherens junction. Interestingly, the authors<sup>1</sup> demonstrate that apical localization of Bitesize also depends on a molecule called PI(4,5)P<sub>2</sub>. This phospholipid is a product of PTEN, a lipid-modifying enzyme that also binds to Baz/Par-3 (ref. 9). So it would seem that the Baz/Par-3 complex might control Bitesize localization by recruiting PTEN to the apical domain, although this has yet to be tested.

Why is the adherens junction unstable in *bitesize* mutants? Interactions with the underlying actin cytoskeleton are thought to stabilize the adherens junction, although the underlying mechanism is not well understood<sup>10</sup>. Pilot *et al.*<sup>1</sup> found a defect in apical actin assembly that precedes disassembly of the adherens junction, suggesting a previously unknown mechanism for stabilizing apical actin in the early embryo. To solve the mystery of how Bitesize regulates actin, they examined data from yet another genome-based approach



**Figure 1 | Assembling adherens junctions between epithelial cells.** Pilot *et al.*<sup>1</sup> show that this complex helps to localize the Bitesize protein to the apical membrane, and that this process is dependent on binding to PI(4,5)P<sub>2</sub>, a phospholipid that is produced by the enzyme PTEN. PTEN also binds to the Baz/Par-3 complex. **b**, Having been localized to the membrane, Bitesize binds to Moesin, a cytoplasmic protein that mediates membrane–cytoskeleton interactions at the apical membrane. **c**, The Moesin–Bitesize interaction organizes the actin filaments of the cytoskeleton, which in turn stabilize the E-cadherin proteins that make up the cell–cell contacts in adherens junctions (AJ). (Adapted from ref. 1.)

\*This article and the paper concerned<sup>1</sup> were published online on 9 July 2006.

that searches for functional protein–protein interactions in yeast. A previous study<sup>11</sup> using this method showed that Bitesize binds to Moesin, a cytoplasmic protein believed to mediate membrane–cytoskeletal interactions in the apical domain<sup>12</sup>.

Pilot *et al.* confirmed this interaction and showed that, in the absence of Bitesize, Moesin does not localize normally in the adherens junction. In addition, they found that knocking down Moesin produced features, including disruption of E-cadherin and epithelial integrity, that were indistinguishable from those of *bitesize* mutants. These results suggest that adherens junctions assemble normally without Bitesize, but cannot stabilize because there is no underlying band of actin microfilaments.

Together, these results fill a large gap in our understanding of how the Baz/Par-3 polarity complex, the actin cytoskeleton and the adherens junction cooperate to establish apical/basal polarity in developing epithelia. However, as with most advances, the study raises as many questions as it answers. Although the results imply that Moesin functions to link the plasma membrane structurally to the underlying cytoskeleton, other studies in fruitflies indicate instead that Moesin regulates a small GTPase called Rho, which in turn regulates cytoskeletal assembly<sup>13,14</sup>. Is it possible that Rho function is spatially regulated in epithelial cells to promote adherens-junction assembly and therefore overall apical/basal polarity? Does the role of synaptotagmin-like proteins in other contexts suggest that vesicle trafficking is involved in adherens-junction formation and apical/basal polarity? Is Bitesize function required in other fruitfly epithelia, and are its mammalian relatives required in mammalian epithelia? Time, and the continued use of genetic and genomic approaches in model systems, will tell. ■

Richard Fehon is in the Department of Molecular Genetics and Cell Biology, 920 East 58th Street, University of Chicago, Chicago, Illinois 60637, USA.  
e-mail: rfehon@uchicago.edu

## ASTRONOMY

## A dwarf-eats-dwarf world

James Liebert

**A white dwarf burnt-out star and a brown dwarf wannabe star have been found in mutual orbit. This fascinating system has had a turbulent past, and its future evolution could be just as spectacular.**

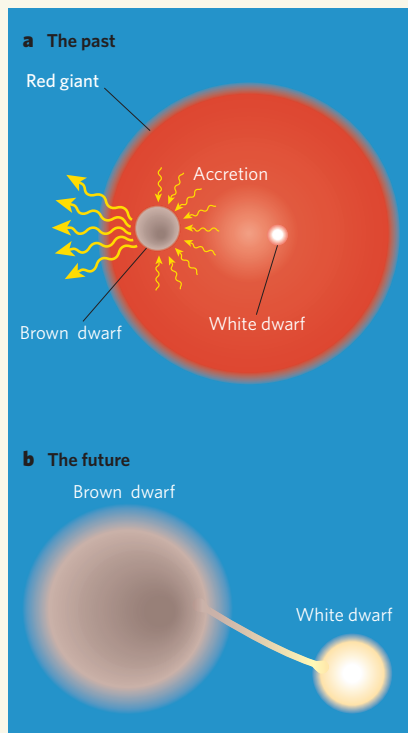
On page 543 of this issue, Maxted *et al.*<sup>1</sup> report observations of a detached astronomical binary system of small orbital separation, known as WD 0137–349. These observations are remarkable for what they imply about the pair's history, and what we can infer from them of its future: within a few billion years, the system will become a so-called cataclysmic variable, in which one star grabs mass from the other. That confirms theoretical predictions of the existence of a new sub-class of these volatile systems.

A detached binary is so-called because its two constituents do not exchange mass as they orbit around their combined centre of mass. In the case of WD 0137–349, the larger (in mass) of the two components is a star that has exhausted its nuclear fuel, a white dwarf. Unusually, the smaller-mass object is a brown dwarf — a would-be star with a mass too low to sustain enough thermonuclear fusion of hydrogen to support itself against its own gravity. By analysing spectra covering the entire 116-minute orbital period of the WD 0137–349 binary, the authors found that the mass of the brown dwarf is just one-twentieth the mass of the Sun — two-thirds of that needed to sustain hydrogen burning.

The significance of this finding is that a body of substellar mass, such as a brown dwarf, can survive the evolutionary process that created the white dwarf. A white dwarf starts out as a normal star, such as our Sun currently, that burns hydrogen to helium in its core. When the supply of hydrogen in the core is exhausted, the star begins to burn its large, dilute surrounding envelope of unburnt hydrogen: it expands rapidly and becomes a 'red giant' centred on a helium core. At this point in the evolution of the WD 0137–349 system, the companion brown dwarf was close enough that it became engulfed in a common envelope surrounding the two stellar cores (Fig. 1a).

Given its low self-gravity, one might expect that the brown dwarf would not survive being swallowed — it would simply be digested, and its nuclear fuel dissipated. But actually, when the red giant shed its envelope, the brown dwarf was spat out again unharmed, accompanied by the spent shadow of the original primary star. Maxted *et al.* demonstrate<sup>1</sup> that the brown dwarf might even have profited from its ordeal: it could have accreted some of the hot material of the common envelope. To test this thesis, the authors emphasize that accurate follow-up observations at infrared wavelengths are needed to test how much the brown dwarf has cooled off.

So what will happen to the system next? Will this white dwarf with its brown dwarf companion continue to display highly unusual, even unique properties? As the authors point out<sup>1</sup>, gravitational waves radiate away from two self-gravitating objects orbiting each other, carrying away angular momentum. The binary orbit grows smaller, and its period decreases. Calculations show that, in 1.4 billion years'



**Figure 1 | Degenerate dwarfs.** **a**, When a star exhausts the hydrogen fuel in its core, it can turn to the hydrogen in its diffuse surrounding envelope. The star expands hugely to become a red giant, centred on a helium core that is the precursor of a white dwarf. In the WD 0137–349 system studied by Maxted *et al.*<sup>1</sup>, in doing so the star swallowed a companion brown dwarf. The companion emerged unharmed, and may indeed have gained mass from the common envelope, causing mass to be lost from the common envelope into space on the outside. **b**, Once all the material of the common envelope has dispersed, the concentrated gravitational field of the white dwarf can suck in material from the now larger brown dwarf, forming a so-called cataclysmic variable.

1. Pilot, F., Philippe, J.-M., Lemmers, C. & Lecuit, T. *Nature* **442**, 580–584 (2006).
2. Tepass, U., Tanentzapf, G., Ward, R. & Fehon, R. *Annu. Rev. Genet.* **35**, 747–784 (2001).
3. Bilder, D., Schober, M. & Perrimon, N. *Nature Cell Biol.* **5**, 53–58 (2003).
4. Tanentzapf, G. & Tepass, U. *Nature Cell Biol.* **5**, 46–52 (2003).
5. Harris, T. J. & Peifer, M. *J. Cell Biol.* **167**, 135–147 (2004).
6. Harris, T. J. & Peifer, M. *J. Cell Biol.* **170**, 813–823 (2005).
7. Serano, J. & Rubin, G. M. *Proc. Natl Acad. Sci. USA* **100**, 13368–13373 (2003).
8. Fukuda, M. *J. Biochem. (Tokyo)* **137**, 9–16 (2005).
9. von Stein, W., Ramrath, A., Grimm, A., Müller-Borg, M. & Wodarz, A. *Development* **132**, 1675–1686 (2005).
10. Gates, J. & Peifer, M. *Cell* **123**, 769–772 (2005).
11. Formstecher, E. *et al.* *Genome Res.* **15**, 376–384 (2005).
12. Bretscher, A., Edwards, K. & Fehon, R. G. *Nature Rev. Mol. Cell Biol.* **3**, 586–599 (2002).
13. Speck, O., Hughes, S. C., Noren, N. K., Kulikaukas, R. M. & Fehon, R. G. *Nature* **421**, 83–87 (2003).
14. Hipfner, D. R., Keller, N. & Cohen, S. M. *Genes Dev.* **18**, 2243–2248 (2004).



time, the orbital period of WD 0137–349 will have declined from its current near 2 hours to just 60–80 minutes. The twosome will then be close enough that the primary star's stronger gravitational field will start grabbing material from the facing side of the brown dwarf (Fig. 1b).

White dwarf stars that accrete matter in this way from a close companion are called cataclysmic variables. The accretion of material rich in hydrogen fuel from the secondary body in such systems leads, eventually, to a violent explosion, known as a nova. In this spectacular thermonuclear runaway event, the brightness of the primary star can increase by a factor of a million or more.

Most known cataclysmic variables, however, are observed in various phases of quiet accretion. Cataclysmic variables are generally divided into long-period and short-period systems, separated by a paucity of systems with periods between 2 and 3 hours (ref. 2). A few systems are found with periods of less than 1 hour: these correspond to systems containing two white dwarfs, which are a special case.

The period gap between 2 and 3 hours is explained by the standard model for cataclysmic-variable evolution<sup>3</sup>. According to this model, most secondaries enter the mass-transfer phase at long orbital periods. Such a secondary has a bloated form, because loss of mass and angular momentum — through convective processes driven by magnetic fields — causes its radius to expand significantly beyond that expected for an isolated star. This state persists as more and more angular momentum is sucked out of the system until the interior of the secondary becomes completely convective near a period of 3 hours.

At this point, the magnetic field weakens, and the secondary star shrinks. In this more compact configuration, the primary body can no longer grab mass from the secondary. Mass transfer does not resume until the period of the system has decreased to around 2 hours, at which point the distance between the primary star and even a non-bloated secondary body is small enough that the primary body can resume picking off matter. Thus, according to the standard model, a single cataclysmic variable evolves from long periods all the way down to the minimum period for normal cataclysmic variables of around 75–80 minutes.

In recent years, however, theoretical evidence has amassed to indicate that a system can enter the cataclysmic-variable state at short periods without going through the longer-period phases. This requires the companion object to have a low mass. If the system first reaches the cataclysmic-variable state near the minimum period, the companion must be very light indeed — a brown dwarf<sup>4–6</sup>. The existence of one cataclysmic variable, named EF Eri, with a period near 80 minutes and a secondary mass very similar to that of the brown dwarf in WD 0137–349, had already been confirmed observationally<sup>7</sup>. But the

orbital period at which EF Eri began its mass-transfer phase is unknown.

Thus Maxted and colleagues' discovery<sup>1</sup> can be put in its proper context. This is the first system to be found that, because of its special history, will definitely enter the mass-transfer state near the minimum period observed for the whole set of known cataclysmic variables. (Another, less clear-cut case has also been recently found<sup>8</sup>). Even a puny brown dwarf can survive being swallowed by a red giant; that observation is bolstering our confidence in the belief that a new class of cataclysmic variables is emerging. ■

James Liebert is at the Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA.  
e-mail: liebert@as.arizona.edu

1. Maxted, P. F. L., Napiwotzki, R., Dobbie, P. D. & Burleigh, M. *R. Nature* **442**, 543–545 (2006).
2. Ritter, H. & Kolb, U. *Astron. Astrophys.* **404**, 301–303 (2003).
3. King, A. R. Q. *J. R. Astron. Soc.* **29**, 1–25 (1988).
4. Kolb, U. & Baraffe, I. *Mon. Not. R. Astron. Soc.* **309**, 1034–1042 (1999).
5. Howell, S. B., Nelson, L. A. & Rappaport, S. *Astrophys. J.* **550**, 897–918 (2001).
6. Politano, M. *Astrophys. J.* **604**, 817–826 (2004).
7. Howell, S. B. *et al. Astrophys. J.* (in the press).
8. Schmidt, G. D. *et al. Astrophys. J.* **630**, L173–L176 (2005).

## ENERGY TECHNOLOGY

# Hydrogen quick and clean

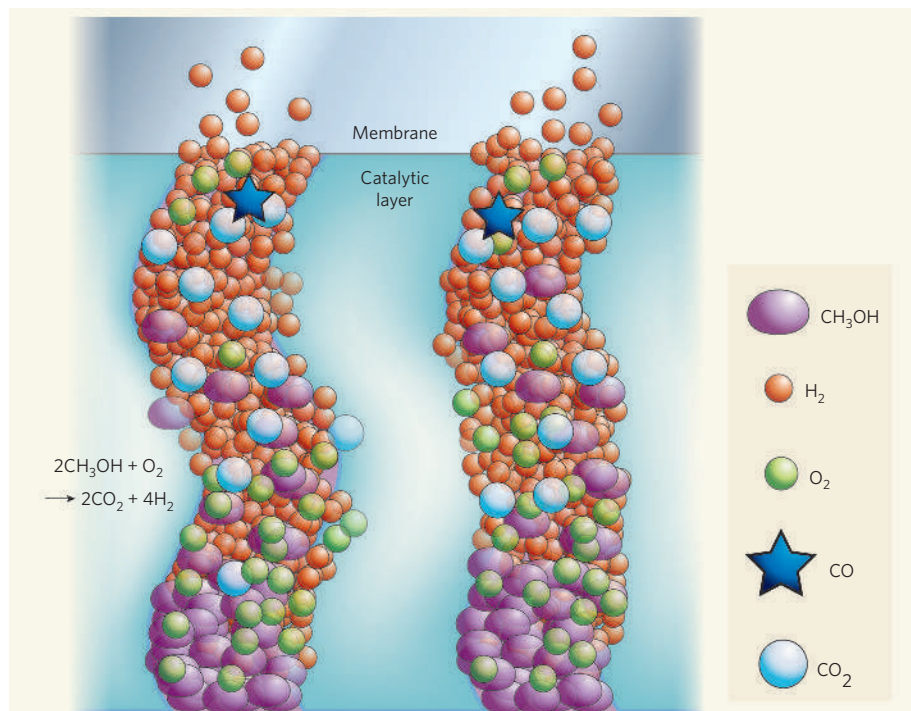
Rich Masel

**Systems for producing pure hydrogen for fuel cells from methanol run into problems with energy efficiency and short lifetimes. Unless, that is, you combine the right catalyst and the right purification membrane.**

While the power demands of portable electronic gadgets — laptops, mobile telephones, iPods and the like — have exploded in the past few years, the storage capacity of the batteries used to power them has not kept pace. In a worldwide push to find alternative power sources, systems based on fuel-cell technologies, with their potentially much higher

energy-storage densities, have been receiving considerable attention.

So far, however, that attention has not translated into practical systems. Writing in *Advanced Materials*, Benjamin Wilhite and colleagues<sup>1</sup> report a substantial advance towards that goal. They have developed a catalytic system that allows them to produce pure



**Figure 1 | All in one.** Hydrogen to power a fuel cell can be produced by methanol oxidation, but needs to be passed through a palladium-based membrane to ensure ultra-purity. Contact with the methanol, or with carbon monoxide produced in the side-reaction  $\text{CH}_3\text{OH} \rightarrow \text{CO} + 2\text{H}_2$ , could corrode such a membrane. Wilhite *et al.*<sup>1</sup> avoid this by covering the membrane with a protective catalytic layer, ensuring that the methanol and other corrosive gases are oxidized before they reach the membrane.

hydrogen, the basic power source of proton-exchange membrane fuel cells, from methanol at a much higher rate per unit volume than has been achieved before.

In the way they work, fuel cells are much like batteries: they use electrochemical reactions to produce electricity. First, a reaction on an anode coated with a catalyst such as platinum produces electrons. These electrons pass through a circuit, delivering power to a device. They then return to the fuel cell, where they are collected by a second reaction on a catalytic cathode, generally producing water as a waste product. For example, in proton-exchange membrane fuel cells, the reactions are  $2\text{H}_2 \rightarrow 4\text{H}^+ + 4\text{e}^-$  (anode) and  $4\text{H}^+ + 4\text{e}^- + \text{O}_2 \rightarrow 2\text{H}_2\text{O}$  (cathode).

Because of its lightness and the large amounts of heat energy it releases on combustion, hydrogen can supply much more energy per unit weight than can the lithium batteries conventionally used in laptops and similar electronic devices. Thus hydrogen fuel cells are an attractive proposition to supplement, or even replace, such batteries.

The catch is that it is difficult to find a way to store or produce enough pure hydrogen in the laptop. One could imagine putting a tank of very pure hydrogen in the laptop and refilling it at filling stations, rather like those proposed for cars. But in practice, the explosive nature of hydrogen rules that possibility out. One could not, for example, countenance taking a laptop with a hydrogen cylinder on to an aeroplane. In fact, of possible fuels, so far only formic acid and methanol have been approved by the dangerous-goods committee of the International Civil Aviation Organization for use in aeroplanes and storage in luggage<sup>2</sup>. Small butane cylinders are allowed in the passenger compartment, but not in the hold. Hydrogen cylinders, metal hydrides and borohydrides are currently banned. A hydrogen-powered laptop, therefore, would need to generate its hydrogen internally from methanol or formic acid.

There have been many suggestions<sup>3–5</sup> for how to generate hydrogen from methanol, which has the chemical formula  $\text{CH}_3\text{OH}$ . A common approach is to perform 'steam reforming' via the reaction  $\text{CH}_3\text{OH} + \text{H}_2\text{O} \rightarrow \text{CO}_2 + 3\text{H}_2$ . The catch here is that a side-reaction yields carbon monoxide:  $\text{CH}_3\text{OH} \rightarrow \text{CO} + 2\text{H}_2$ . Carbon monoxide is highly poisonous for the catalysts used in fuel cells: if as much as one part per million of carbon monoxide is mixed in with the hydrogen that reaches the anode, this will, over a few weeks of continuous operation, poison all the active sites on the catalyst and effectively kill the fuel cell. A trace presence of carbon dioxide can also be fatal, as this produces carbon monoxide through the reverse water–gas shift reaction:  $\text{CO}_2 + \text{H}_2 \rightarrow \text{CO} + \text{H}_2\text{O}$ .

In the past, differentially permeable membranes have been used to purify hydrogen of such trace gases, with limited success. Membranes made from the lattice-structured mineral zeolite or mesoporous silica allow too

much carbon monoxide through for practical use. Palladium and palladium–silver membranes produce pure hydrogen, but so far have been found to degrade quickly in the presence of carbon monoxide and methanol. A further problem is that it takes considerable energy to push hydrogen through a membrane that is of the thickness required for mechanical stability. Thus, no one has yet been able to produce hydrogen of sufficient purity at a high enough rate to power a laptop for an extended period.

Wilhite *et al.*<sup>1</sup> found that, by building a composite structure that combined a hydrogen-permeable palladium–silver membrane with a catalyst for the hydrogen-producing methanol-oxidation reaction  $2\text{CH}_3\text{OH} + \text{O}_2 \rightarrow 2\text{CO}_2 + 4\text{H}_2$  (Fig. 1), they could prevent the membrane from degrading. Protective layers are often used<sup>6</sup> in commercial applications to protect sensitive materials from corrosive gases by forming an impenetrable barrier in front of them. But such an impermeable system is of no use in a fuel-cell system, where hydrogen must pass through both the protective layer and the membrane to reach the anode of the fuel cell.

But Wilhite and colleagues use a permeable catalytic layer that assists in the oxidation of most carbon monoxide and methanol, rendering them harmless before they reach the membrane, while letting hydrogen past. Any traces of the corrosive gases that reach the membrane are small enough to be oxidized there before they can do any damage, leaving the hydrogen to filter through. Thus, hydrogen can be both

produced and purified in one stage. That is important for portable devices, but this all-in-one approach to corrosion protection could also be useful in many other chemically reacting systems.

As with most scientific endeavour, progress comes in small steps. Wilhite *et al.* have demonstrated<sup>1</sup> that their membrane is stable for days, as opposed to a few hours with an unprotected membrane. But a laptop's power supply needs to be stable for years. More thermal management is needed, too: the temperature in the authors' system is about 400 °C, higher than you would be comfortable with in your lap or pocket. Nevertheless, this step represents a significant advance towards practical portable fuel-cell power systems. ■

Rich Masel is in the Departments of Chemical and Biomolecular Engineering, and Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 600 South Mathews, Urbana, Illinois 61801, USA.  
e-mail: r-masel@uiuc.edu

1. Wilhite, B. A., Weiss, S. E., Ying, J. Y., Schmidt, M. A. & Jensen, K. F. *Adv. Mater.* **18**, 1701–1704 (2006).
2. International Civil Aviation Organization Report AN-WP/8084 (18 January 2006).
3. Holladay, J. D., Wang, Y. & Jones, E. *Chem. Rev.* **104**, 4767–4789 (2004).
4. Masel, R. I., Gold, S. & Zheng, N. in *Encyclopedia of Chemical Processing* (ed. Lee, S.) 1643–1661 (Taylor & Francis, New York, 2005).
5. Ferreira-Aparicio, P., Benito, M. J. & Sanz, J. L. *Catal. Rev. Sci. Eng.* **47**, 491–588 (2005).
6. Khaladkar, P. R., Hare, C. H. & Norsworthy, R. in *Uhlig's Corrosion Handbook* (ed. Revie, R. W.) 965–1069 (Wiley, New York, 2000).

## SUPERCONDUCTIVITY

# Hot vibes

Alex de Lozanne

**Lattice vibrations — phonons — have long been implicated in conventional low-temperature superconductivity. That they could also have a supporting role when the heat is turned up had been dismissed.**

Just like Peter Sellers' hapless Hindu actor in the film *The Party*, the idea that vibrations of a solid's crystal lattice might be crucial in high-temperature superconductivity has been shot down a dozen times, only to rise defiantly again and again. This time, as Lee *et al.* report in this issue (page 546)<sup>1</sup>, our battered old friend might stay up, with powerful support from another quarter.

In classic superconductors such as lead or mercury, electrons flow without experiencing resistance, but only at very low temperatures. Here, crystal vibrations — known as phonons — provide an attractive mechanism that pairs up electrons possessing opposite senses of rotation, or spins. As the net spin of such an electron pair is zero, it is not subject to the Pauli exclusion principle, so it can share the same quantum state with an unlimited number of

other pairs. Thus, all the pairs participate in a perfectly coherent quantum dance.

The energy needed to break a pair out of this superconducting dance is called the energy gap, and both the dance circle and the energy gap become larger as more and more electron pairs enter the ballroom. The unpaired electrons pile up at the edge of the dancing circle, as if eager to enter the dance floor. Phonons do not bother the electrons once coupled, but they constantly pull all unpaired electrons away from the centre, so the pile at the edge of the dance floor forms a bulge at a fixed distance from it (Fig. 1).

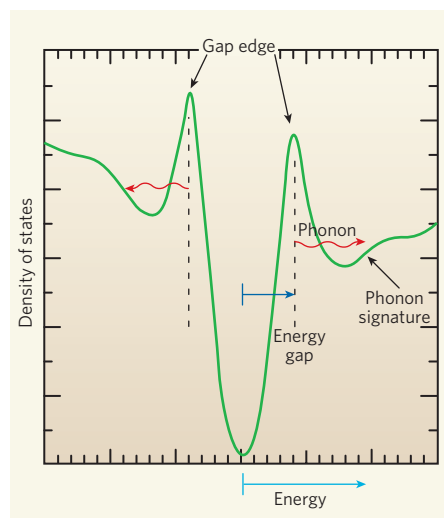
It is this bulge of unpartnered electrons that Lee and colleagues have now spotted<sup>1</sup> in  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ , a cuprate (copper-oxide) superconductor that ceases to be superconducting only above a high transition



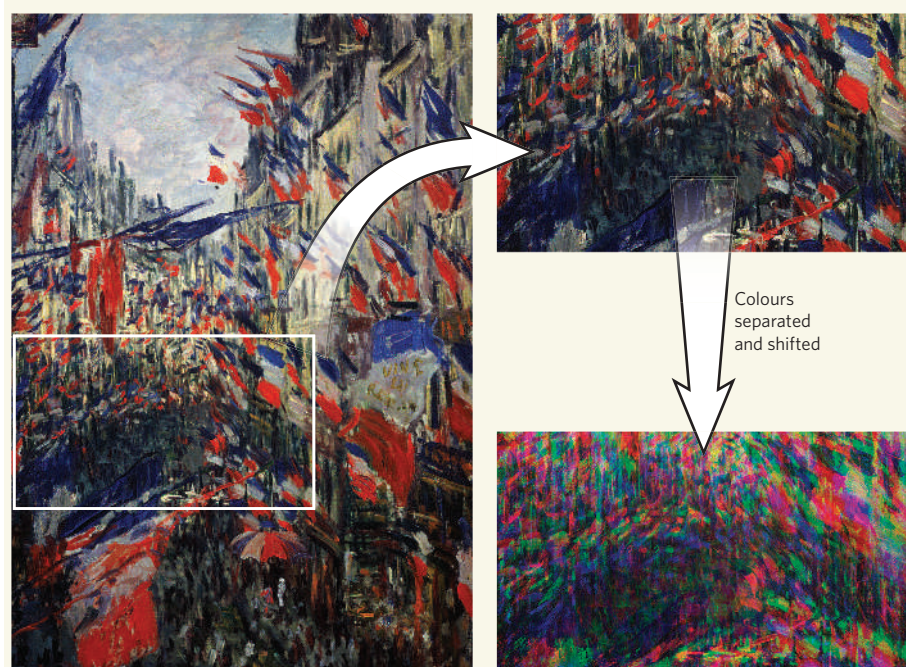
temperature. The authors used the technique of scanning tunnelling spectroscopy, in which voltage measurements from a conducting probe provide a precise picture of the density of energy states on an atomic scale. They could thus measure the distance from the edge of the dancing circle to the bulge, which is also a direct measure of the energy of the phonons pulling the unpartnered electrons away.

Lee and colleagues' analysis is simplified by the fact that phonons are quantized — they are allowed only specific discrete energy values — and that there are fewer phonons around at the low temperatures under consideration. The authors performed the ultimate test to confirm the phononic origin of the effect by growing their  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$  superconductor in an environment rich in the heavy oxygen isotope  $^{18}\text{O}$ , rather than the normal  $^{16}\text{O}$ . The resulting increase in the mass of the material's lattice reduces the frequency of its oscillation, as in any simple oscillator, and thus the allowed phonon energies. And, indeed, the phonon peaks in the tunnelling spectra shifted exactly as expected.

These observations do not necessarily mean that phonons are initially responsible for making superconducting pairs in high-temperature superconductors<sup>2</sup>, as they are known to be in low-temperature superconductors. Lee *et al.*<sup>1</sup> do provide circumstantial evidence that this is the case: they find superconductivity is strongest in places where the phonon signal is strongest. Unfortunately, this would also be true, to a lesser extent, if the phonons were not involved in pairing, because the strengthening of superconductivity would cause a sharpening of the peaks at the energy gap that would enhance the signatures of all phonons. This particular issue might be resolved by a more



**Figure 1 | Leading the quantum dance.** As superconductivity develops in a material, it clears a gap in the density of energy states surrounded by two sharp peaks. Phonons constantly pull unpaired electrons away from this 'dance circle', producing characteristic bulges at a fixed distance from the peaks — the signature spotted by Lee and colleagues<sup>1</sup>.



**Figure 2 | Blurred impressions.** When the red, green and blue components of Claude Monet's *Rue St Denis, fête du 30 juin 1878* — representing photons of different energy — are shifted by 10% of the frame size and added back together, the results (bottom right) are even more highly impressionistic than the original. The data of Lee *et al.*<sup>1</sup> indicate that an analogous smearing of phonon energies has concealed them as the enabling force behind high-temperature superconductivity.

sophisticated analysis that normalizes out such effects.

So why did it take more than 20 years of intensive research to get this far? An answer lies in the inhomogeneity of superconductivity as measured by the energy gap. Even the most perfect  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$  crystals show energy-gap values that change by a factor of three over a distance of nanometres. If phonons were the mechanism responsible, the phonon peaks would shift with the gap peaks, so any technique that averages over more than a nanometre will produce a blurred average phonon energy. To see the confusing effect this scrambling of information in energy and space can have, one need only consider the analogous effect with photons (Fig. 2).

The discovery of cuprate superconductors<sup>3</sup> with transition temperatures up to 95 K — and now up to 138 K (ref. 4) — also seemed to rule out phonons as a universal mechanism to explain superconductivity at all temperatures<sup>5</sup>. The stronger the electron–phonon interaction, the higher the transition temperature of a superconductor to a non-superconducting state must be. But one cannot simply go on increasing the strength of the electron–phonon interaction, as the crystal structure of the material would experience so great an internal tension that it would collapse. Such calculations admittedly do not take into account the fact that most high-temperature superconductors have a complex, layered structure, and that defects in the crystal could possibly have stabilizing effects.

Another argument against a phononic origin for high-temperature superconductivity is

that the quantum-mechanical wavefunction describing a superconducting state, when traced in momentum ( $k$ ) space, resembles a four-leaf clover or shamrock, whereas phonons are predicted to produce a circular pattern. But how well such predictions work when the electron–phonon interaction is strong and the phonons are very unevenly distributed around the crystal is unclear.

Intriguingly, phonons are also being fingered in explaining the properties of manganites<sup>6</sup>, a family of compounds that have similar crystal structure to the cuprate superconductors, but with spectacular magnetic and resistive properties instead.

The results presented by Lee *et al.* are the culmination of years of coordinated work by collaborating research groups in tunnelling spectroscopy, crystal growth and characterization, and theory. There is, of course, always room for alternative explanations, but the clear message is that phonons cannot be ejected from the party hosted by high-temperature superconductivity. They might indeed be the ones making the whole fling possible.

Alex de Lozanne is in the Department of Physics, University of Texas at Austin, 1 University Station C1600, Austin, Texas 78712-0264, USA.  
e-mail: delozanne@physics.utexas.edu

1. Lee, J. *et al.* *Nature* **442**, 546–550 (2006).
2. Pilgram, S., Rice, T. M. & Sigrist, M. preprint available at <http://arxiv.org/abs/cond-mat/0605646> (2006).
3. Bednorz, J. G. & Müller, K. A. *Z. Phys. B* **64**, 189–193 (1986).
4. Bonn, D. A. *Nature Phys.* **2**, 159–168 (2006).
5. Varma, C. M., Schmitt-Rink, S. & Abrahams, E. *Solid State Commun.* **62**, 681–685 (1987).
6. Mannella, N. *et al.* *Nature* **438**, 474–478 (2005).

## BRIEF COMMUNICATIONS

## Bees associate warmth with floral colour

Pollinators may be seeking more than just food as a reward when they choose one flower over another.

Floral colour signals are used by pollinators as predictors of nutritional rewards, such as nectar<sup>1–3</sup>. But as insect pollinators often need to invest energy to maintain their body temperature<sup>4</sup> above the ambient temperature, floral heat might also be perceived as a reward. Here we show that bumblebees (*Bombus terrestris*) prefer to visit warmer flowers and that they can learn to use colour to predict floral temperature before landing. In what could be a widespread floral adaptation, plants may modulate their temperature to encourage pollinators to visit.

Some beetles spend extended periods (about 24 hours) inside specialized thermogenic flowers, even in the absence of a nutritional reward<sup>5</sup>, and basking insects will take advantage of floral suntraps<sup>6</sup>. Visits to flowers by pollinating insects in order to imbibe carbohydrates in nectar are typically much briefer. But it is possible that endothermic pollinators might also seek a metabolic reward in the form of heat, given that the temperature of floral nectar is the same as the flower containing it. Differences in floral temperature occur widely between and within plant species<sup>6–9</sup> and, if these variations can influence the preference of pollinators,

pollinators may forage adaptively by paying attention to temperature when choosing between flowers<sup>10</sup>.

To test whether warmer nectar is preferred by pollinators, we connected a bumblebee nest box to a flight arena where sucrose solution (20% by volume) was available from two identical feeders, one at room temperature (18.5 °C ± 0.3 s.d.) and the other at 18.5 °C, 22.5 °C, 27 °C or 29.5 °C (for details, see supplementary information). There was a significant increase in bees' preference for the warmer feeder (Pearson's  $R = 0.9870$ ,  $P = 0.012$ ), and this preference was significant when the temperature difference was 4 °C or more (Fig. 1a). In this case, bees were using spatial positioning to identify the warmer feeder.

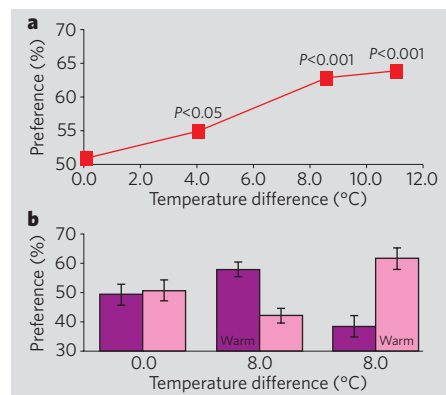
To test whether bees can learn to use flower colour to identify warmer flowers, we exposed them to coloured artificial flowers (four purple 'flowers' at 28.8 ± 0.2 °C and four pink 'flowers' at 20.8 ± 0.1 °C), which were positioned randomly and which each presented 20 µl sucrose solution (40% by volume) (see supplementary information).

Choice frequency for bees landing on the warmer purple flowers was 58.0% (± 2.6 s.d.;  $\chi^2 = 25.6$ , d.f. = 1,  $P < 0.001$ ;  $n = 10$  bees) (Fig. 1b). This was significantly higher than the choice frequency in a control group, for which there was no temperature difference between the purple and pink flowers, and indicates that the bees did not simply prefer purple flowers (Fig. 1b; mean, 49.4% ± 3.0 s.d.; independent samples  $t$ -test,  $t = 9.54$ , d.f. = 18,  $P < 0.001$ ;  $n = 10$  bees per treatment). When the pink flowers were warmer than the purple ones, the pink colour was preferred (Fig. 1b; mean, 61.6% ± 3.8 s.d.;  $\chi^2 = 53.8$ , d.f. = 1,  $P < 0.001$ ;  $n = 10$  bees).

We conclude that the bees preferred to land on the warmer flowers, even though the similarly coloured alternative contained the same nutritional reward. In another control experiment in which flowers varied in temperature but not colour, discrimination fell to chance levels (50.8% ± 3.1 s.d.;  $\chi^2 = 0.3$ , d.f. = 1;  $P = 0.64$ , NS;  $n = 10$  bees), indicating that the bees must use colour as a cue, rather



Hot spot: a thermographic image of a bumblebee taken with an infrared camera — brightness indicates higher temperature.



**Figure 1 | Temperature preferences and flower-colour use by bees.** **a**, Bee preference for sucrose solution at different temperatures above room temperature (18.5 °C);  $P$  values for  $\chi^2$ ; d.f. = 1. **b**, Bees learn to associate colour with temperature: when purple flowers are warmer than pink flowers by 8 °C, bees prefer the warmer purple flowers (central column pair); when pink flowers are warmer, these are chosen more often (right column pair). When there is no temperature difference, bees show no preference for either colour (left column pair). For each bee 100 choices were evaluated, and ten bees were used per group.

than directly gauging temperature by remote perception.

Our findings indicate that floral temperature can serve as an additional reward for pollinator insects when nutritional rewards are also available. They may also have implications for the evolution of specific floral structures and for the connection between floral sensory cues, floral temperature and pollinator behaviour<sup>9</sup>.

**Adrian G. Dyer\***, **Heather M. Whitney\***,  
**Sarah E. J. Arnold\***, **Beverley J. Glover\***,  
**Lars Chittka†**

\*Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK

†Biological and Chemical Sciences, Queen Mary College, University of London, London E1 4NS, UK  
e-mail: l.chittka@qmul.ac.uk

1. Menzel, R. & Backhaus, W. In *The Perception of Colour* (ed. Gouras, P.) 262–293 (Macmillan, London, 1991).
2. Chittka, L., Dyer, A. G., Bock, F. & Dornhaus, A. *Nature* **424**, 388 (2003).
3. Giurfa, M. *Naturwissenschaften* **91**, 228–231 (2004).
4. Heinrich, B. & Esch, H. *Am. Sci.* **82**, 164–170 (1994).
5. Seymour, R. S., White, C. R. & Gibernau, M. *Nature* **426**, 243–244 (2003).
6. Kevan, P. G. *Science* **189**, 723–726 (1975).
7. Sapir, Y., Shmida, A. & Ne'eman, G. *Oecologia* **147**, 53–59 (2006).
8. Miller, G. A. *Oecologia* **70**, 155–160 (1986).
9. Comba, L. et al. *Plant Cell Environ.* **23**, 639–647 (2000).
10. Thien, L. B., Azuma, H. & Kawano, S. *Int. J. Plant Sci.* **161** (suppl.), 225–235 (2000).

Supplementary information accompanies this communication on Nature's website.

Received 6 April; accepted 30 June 2006.

Competing financial interests: declared none.  
doi:10.1038/442525a



## SUPERHYDROPHOBICITY

## Drying transition of confined water

Long-range hydrophobic interactions operating underwater are important in the mediation of many natural and synthetic phenomena, such as protein folding, adhesion and colloid stability. Here we show that rough hydrophobic surfaces can experience attractive forces over distances more than 30 times greater than any reported previously, owing to the spontaneous evaporation of the intervening, confined water. Our finding highlights the importance of sur-

face roughness in the interaction of extended structures in water, which has so far been largely overlooked.

The existence of 'long-range' hydrophobic interactions has been debated for more than 25 years<sup>1–5</sup>, because their reported range of 1–100 nanometres exceeds that of van der Waals forces and cannot be explained by water restructuring. However, investigations have been limited to smooth, flat model surfaces<sup>4</sup>, even though most surfaces are rough. Roughness strongly influences wetting — as evidenced by the high water-contact angles ( $\theta \geq 160^\circ$ ) and rolling of water droplets on lotus leaves<sup>6</sup>.

Knowing that superhydrophobic surfaces occur naturally and that living systems operate mainly in water, we investigated the interaction of rough superhydrophobic surfaces beneath the water surface. Interfacial-force microscopy<sup>7</sup> and optical imaging were used to examine the interaction between two approaching or retracting superhydrophobic surfaces (water contact angle  $\theta$  of about  $170^\circ$ ; see supplementary information) submerged in either air-equilibrated or de-aerated water.

We discovered a very-long-range hydrophobic interaction that was due to out-of-contact evaporation, or 'cavitation', of the intervening water at tip-to-substrate separations ranging from 0.8  $\mu\text{m}$  to as much as 3.5  $\mu\text{m}$ . Cavitation is a first-order phase transition characterized by a sudden, strong attractive force (Fig. 1a) and by the appearance of a vapour bridge spanning the tip-to-substrate gap (Fig. 1b–e, and see supplementary information for details).

Pre-existing 'nanobubbles' of a size commensurate with the interaction length have been proposed as a source of long-range interactions<sup>8</sup>. We therefore used *in situ* confocal imaging to search for bubbles on an isolated, flat superhydrophobic surface (for details, see supplementary information). Neither the microscopy nor neutron reflectivity experiments<sup>9</sup> provide evidence for bubbles — certainly not micrometre-sized bubbles. We therefore argue that cavitation is a consequence of, and thermodynamically consistent with, the properties of confined water<sup>10</sup>. The critical separation,  $D$ , below which cavitation is thermodynamically favoured can be estimated from Laplace's equation as 1.4  $\mu\text{m}$  (see supplementary information): this is a lower bound and will increase if  $\Delta p$ , the pressure difference across the interface, is reduced by incorporation of air into the cavity.

Unambiguous cavitation has previously been observed only following the contact of smooth hydrophobic surfaces<sup>3,8,11</sup>. Although it is thermodynamically acceptable, out-of-contact cavitation has been modelled as having a large activation barrier<sup>12</sup>, so we need to explain

why we observe it and how kinetic barriers are surmounted.

First, we used interfacial-force microscopy<sup>7</sup>, in which interfacial forces are counterbalanced, to avoid snap-to-contact (when the rate of change of the force exceeds the spring constant and the two surfaces snap together uncontrollably, as occurs in atomic-force microscopy and in surface-force apparatus studies<sup>5</sup>). Second, the calculated  $D$  for submerged superhydrophobic surfaces is more than ten times that of smooth surfaces, which increases the probability of cavity nucleation at larger separations. Third, the submerged superhydrophobic surface presents an intrinsic liquid/vapour/solid interface, which may serve as a heterogeneous nucleation surface.

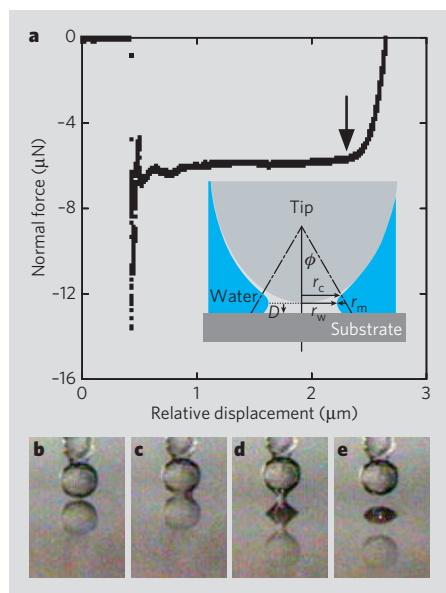
We simulated the molecular dynamics of a model fluid, which showed the necessary liquid–vapour behaviour, confined between two closely spaced surfaces containing opposing ( $\theta = 180^\circ$ ) superhydrophobic patches. The simulations show that cavitation occurs by the growth of capillary-like fluctuations of vapour films extending from the superhydrophobic surfaces, leading to the sudden formation and growth of a vapour bridge.

Our experiments reveal that cavitation can be one source of long-range hydrophobic interactions. For rough surfaces, the interaction length extends to micrometre scales. Correspondingly, interaction lengths of hundreds of nanometres, observed for smooth surfaces (for example, snap-to-contact), might also be explained by cavitation, in agreement with thermodynamic expectations.

Seema Singh\*, Jack Houston\*, Frank van Swol\*†, C. Jeffrey Brinker\*†

\*Sandia National Laboratories, PO Box 5800, Albuquerque, New Mexico 87185, USA  
e-mail: cjbrink@sandia.gov

†University of New Mexico, Albuquerque, New Mexico 87131, USA



**Figure 1 | Cavitation between superhydrophobic surfaces.** **a**, Plot of force versus displacement for the underwater approach of a tip towards a flat surface. The experiment starts at zero micrometres relative displacement (arbitrarily defined). The sudden development of a negative (attractive) force at about 0.4  $\mu\text{m}$  relative displacement corresponds to cavitation. Contact with the surface is indicated by the inflection at about 2.2  $\mu\text{m}$  (arrow). The distance between the onset of adhesion and contact (1.8  $\mu\text{m}$ ) is the distance over which cavitation occurs for this sample. The inset shows the cavitation geometry:  $\Delta p = \gamma(1/r_w - 1/r_m)$ , where  $\Delta p$  is the pressure difference across the interface,  $\gamma$  is the liquid–vapour interfacial tension,  $r_m$  is the radius of the meniscus,  $r_w$  is the radius of its waist, and  $r_c$  is the contact radius of the cavity on the tip surface.  $D$  is the critical separation below which cavitation is thermodynamically favoured. See also supplementary information. **b–e**, Optical images of cavitation: **b**, position of superhydrophobic tip and substrate just before cavitation; **c**, cavitation occurs about 33 ms later; **d**, cavity meniscus, as seen during tip retraction, one frame before its unstable collapse; **e**, a cavity 'bubble' is left behind on both the tip and substrate. These bubbles, attributed to air supplied from water and the porous superhydrophobic surface, are unstable and are readsorbed in about 6 seconds. In all frames, the circular image at the bottom is the reflection of the spherical 150- $\mu\text{m}$ -diameter tip in the flat superhydrophobic surface.

1. Blake, T. D. & Kitchener J. A. *J. Chem. Soc. Farad. Trans.* **168**, 1435–1442 (1972).
2. Christenson, H. K. & Claesson, P. M. *Adv. Colloid Interface Sci.* **91**, 391–436 (2001).
3. Christenson, H. K. & Claesson, P. M. *Science* **239**, 390–392 (1988).
4. Israelachvili, J. & Pashley, R. *Nature* **300**, 341–342 (1982).
5. Stevens, H., Considine, R. F., Drummond, C. J., Hayes, R. A. & Attard, P. *Langmuir* **21**, 6399–6405 (2005).
6. Lafuma, A. & Quéré, D. *Nature Mater.* **2**, 457–460 (2003).
7. Joyce, S. A. & Houston, J. E. *Rev. Sci. Instr.* **62**, 710–715 (1991).
8. Parker, J. L., Claesson, P. M. & Attard, P. *J. Phys. Chem.* **98**, 8468–8480 (1994).
9. Doshi, D. et al. *Langmuir* **21**, 7805–7811 (2005).
10. Lum, K., Chandler, D. & Weeks, J. D. *J. Phys. Chem. B* **103**, 4570–4577 (1999).
11. Pashley, R. M., McGuigan, P. M., Ninham, B. W. & Evans, F. D. *Science* **229**, 1088–1089 (1985).
12. Luzar, A. *J. Phys. Chem. B* **108**, 19859–19866 (2004).

Supplementary information accompanies this communication on Nature's website.

Received 15 December 2005; accepted 15 June 2006.

Competing financial interests: declared none.

doi:10.1038/442526a

BRIEF COMMUNICATIONS ARISING online  
[www.nature.com/bca](http://www.nature.com/bca) see Nature contents.

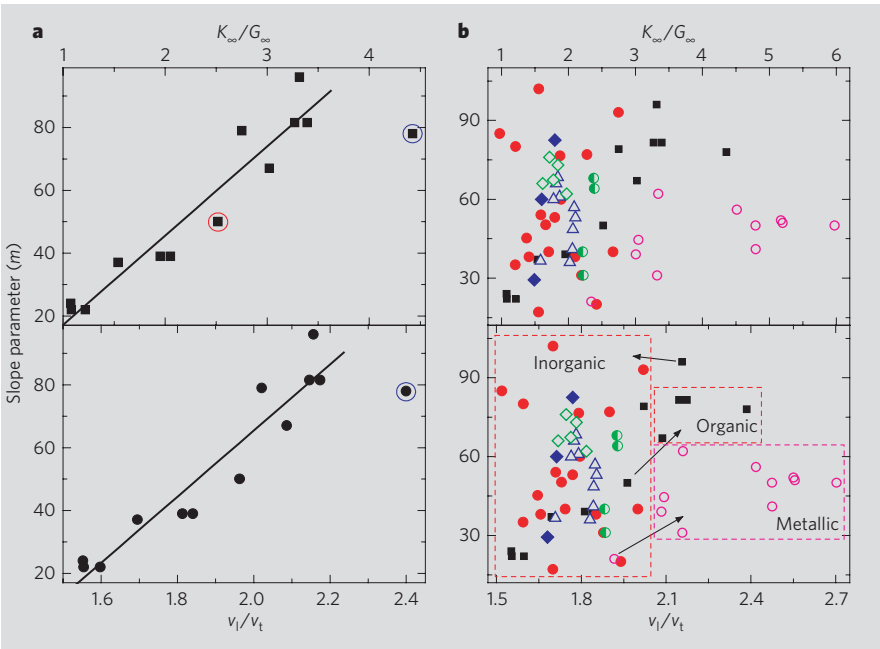
GLASS BEHAVIOUR

Poisson’s ratio and liquid’s fragility

Arising from: V. N. Novikov & A. P. Sokolov *Nature* **431**, 961–963 (2004)

The lack of a reliable theory of glass physics has led to the pursuit of correlations between various glass or viscous liquid parameters, one of which is the slope  $m$  of the plot of  $\log(\text{viscosity})$  against  $T_g/T$ , extrapolated at the glass-transition temperature,  $T_g$ , also termed ‘fragility’. Novikov and Sokolov<sup>1</sup> conclude that the value of  $m$  for a liquid varies linearly with the ratio of the instantaneous bulk and shear moduli,  $K_\infty/G_\infty$ , of its glass according to the relation  $m = 29(K_\infty/G_\infty - 0.41)$ . Because of the obvious importance of the elastic properties of a glass, we have investigated the basis for this relation<sup>1</sup> and find that its premise is flawed because of the unjustifiable preference for an empirical variation of  $m$  with elastic properties, and because of the selected use of glasses. When more glasses are considered in the same way,  $m$  does not seem to be linearly related to  $K_\infty/G_\infty$ .

The square of the ratio of the propagation velocity of longitudinal and transverse ultrasonic (or hypersonic) waves through an isotropic medium,  $(v_l/v_t)^2$ , yields  $K_\infty/G_\infty = [(v_l/v_t)^2 - (4/3)]$ . First, we present the data of the authors’ Fig. 2 in ref. 1 as a plot of  $m$  against  $K_\infty/G_\infty$  and against  $v_l/v_t$  (Fig. 1a), with the incorrectly plotted data point indicated (note that the glasses are binary-component, molecular, hydrogen-bonded and network-structure



**Figure 1 | Relation between slope parameter  $m$  and the elastic properties of glasses.** **a, b**, Plots of  $m$  against  $K_\infty/G_\infty = [(v_l/v_t)^2 - (4/3)]$  and against  $v_l/v_t$ , showing a preference for data fitting (**a**) and a lack of linear relation with  $m$  (**b**). Data in **a** are from Novikov and Sokolov<sup>1</sup> and show the erroneously plotted point for glycerol (red dot) and for  $m$ -toluidine (blue dot); in **b**, black symbols are from Novikov and Sokolov<sup>1</sup> and coloured symbols are data for glasses listed in Table 1. Filled circles, inorganic glasses; open triangles, alkali borates; open diamonds, bismuth borates; filled diamonds, alkali germanates; half-filled circles, arsenic selenides; and open circles, metallic glasses.

**Table 1 | Data set of 50 different types of glass**

| Type of glass                                                                                                                          | References | Type of glass                                                                                   | References |
|----------------------------------------------------------------------------------------------------------------------------------------|------------|-------------------------------------------------------------------------------------------------|------------|
| <b>Inorganic glasses</b>                                                                                                               |            | <b>Metallic glasses</b>                                                                         |            |
| GeS <sub>2</sub>                                                                                                                       | 2, 3       | Pd <sub>77.5</sub> Si <sub>16.5</sub> Cu <sub>6</sub>                                           | 23, 24     |
| NBS710 (70.5SiO <sub>2</sub> 8.7Na <sub>2</sub> O 7.7K <sub>2</sub> O 11.6CaO (percentage by weight)                                   | 4, 5       | Pt <sub>60</sub> Ni <sub>15</sub> P <sub>25</sub>                                               | 23, 24     |
| ZBLA (58ZrF <sub>4</sub> 33BaF <sub>2</sub> 5LaF <sub>3</sub> 4AlF <sub>3</sub> )                                                      | 6, 5       | Zr <sub>41.2</sub> Ti <sub>13.8</sub> Cu <sub>12.5</sub> Ni <sub>10</sub> Be <sub>27.5</sub>    | 25, 26     |
| ZBLAN20 (53ZrF <sub>4</sub> 20BaF <sub>2</sub> 4LaF <sub>3</sub> 3AlF <sub>3</sub> 20NaF)                                              | 6, 5       | P <sub>39</sub> Ni <sub>10</sub> Cu <sub>30</sub> P <sub>21</sub>                               | 25, 26     |
| HLBAN20 (53HfF <sub>4</sub> 20BaF <sub>2</sub> 4LaF <sub>3</sub> 3AlF <sub>3</sub> 20NaF)                                              | 6, 5       | Ce <sub>70</sub> Al <sub>10</sub> Cu <sub>10</sub> Ni <sub>10</sub>                             | 27, 27     |
| 2BiCl <sub>3</sub> -KCl                                                                                                                | 5, 7       | Pd <sub>40</sub> Ni <sub>40</sub> P <sub>20</sub>                                               | 25, 28     |
| As <sub>2</sub> S <sub>3</sub>                                                                                                         | 8, 9       | Zr <sub>46.75</sub> Ti <sub>8.25</sub> Cu <sub>7.5</sub> Ni <sub>10</sub> Be <sub>27.5</sub>    | 25, 29     |
| As <sub>2</sub> Se <sub>3</sub>                                                                                                        | 10, 9      | Pd <sub>64</sub> Ni <sub>16</sub> P <sub>20</sub>                                               | 25, 30     |
| As <sub>x</sub> S <sub>100-x</sub> (x: 5, 10, 20 or 30)                                                                                | 11, 12     | Pd <sub>16</sub> Ni <sub>64</sub> P <sub>20</sub>                                               | 25, 30     |
| 15LiCl-85H <sub>2</sub> O                                                                                                              | 13, 14     | Zr <sub>50</sub> Cu <sub>50</sub>                                                               | 31, 32     |
| Ca(NO <sub>3</sub> ) <sub>2</sub> -8H <sub>2</sub> O                                                                                   | 15, 15     | Pr <sub>60</sub> Cu <sub>20</sub> Ni <sub>10</sub> Al <sub>10</sub>                             | 33, 33     |
| BSC (borosilicate) 70SiO <sub>2</sub> 11B <sub>2</sub> O <sub>3</sub> 9Na <sub>2</sub> O 7K <sub>2</sub> O 3BaO (percentage by weight) | 16, 17     | <b>Alkali-borate glasses</b>                                                                    |            |
| CaAl <sub>2</sub> Si <sub>2</sub> O <sub>8</sub> (anorthite)                                                                           | 4, 18      | xLi <sub>2</sub> O-(100-x)B <sub>2</sub> O <sub>3</sub> (x: 10, 20 or 30)                       | 34, 20     |
| CaMgSiO <sub>6</sub> (diopside)                                                                                                        | 4, 18      | xNa <sub>2</sub> O-(100-x)B <sub>2</sub> O <sub>3</sub> (x: 10, 15, 20, 25, 30, 35 or 40)       | 34, 20     |
| 20K <sub>2</sub> O-80B <sub>2</sub> O <sub>3</sub>                                                                                     | 19, 20     | <b>Bismuth-borate glasses</b>                                                                   |            |
| 45B <sub>2</sub> O <sub>3</sub> -55SiO <sub>2</sub>                                                                                    | 19, 21     | xBi <sub>2</sub> O <sub>3</sub> -(100-x)B <sub>2</sub> O <sub>3</sub> (x: 20, 25, 30, 35 or 45) | 22, 35     |
| 33Li <sub>2</sub> O-67SiO <sub>2</sub>                                                                                                 | 19, 22     |                                                                                                 |            |
| 33Na <sub>2</sub> O-67SiO <sub>2</sub>                                                                                                 | 22, 22     |                                                                                                 |            |
| xNa <sub>2</sub> O-(100-x)GeO <sub>2</sub> (x: 5, 10 or 30)                                                                            | 22, 22     |                                                                                                 |            |

The first reference number is the reference for the  $m$  value, and the second is the reference for the elastic properties. In cases, especially for metallic glasses, where several  $m$  values were reported, we used an average value.



type); straight lines are least-square linear fits to the data. It is evident that the linear relation fits the data equally well irrespective of whether  $m$  is plotted against  $K_{\infty}/G_{\infty} = [(v_l/v_t)^2 - (4/3)]$  or against  $v_l/v_t$ . There seems to be no reason why Novikov and Sokolov<sup>1</sup> preferred a linear relation between  $m$  and  $K_{\infty}/G_{\infty}$  to a linear relation between  $m$  and  $v_l/v_t$ .

Second, for a more detailed investigation of the available data on glasses listed in Table 1, we present the plot of  $m$  against  $K_{\infty}/G_{\infty} = [(v_l/v_t)^2 - (4/3)]$  and against  $v_l/v_t$  in Fig. 1b, with the 13 data points taken from Novikov and Sokolov<sup>1</sup>. It is evident that there is no linear relation between  $m$  and the elastic properties of glasses. The glasses are grouped into three types: inorganic, organic and metallic. The data also show that  $m$  does not linearly increase with  $K_{\infty}/G_{\infty}$ , even within one group of glasses.

Evidently, both a preference for a linear variation of  $m$  with  $K_{\infty}/G_{\infty}$  (or  $v_l^2/v_t^2$ ) over a linear variation of  $m$  with  $v_l/v_t$  and the limited data used in the correlation have led to an unreliable conclusion by Novikov and Sokolov<sup>1</sup>. Therefore, the physical significance of the subsequent discussion of the glass and liquid properties based on this relation<sup>1</sup> is questionable.

Our findings are significant because Poisson's ratio for a glass is important in technology and in structural design. It is a measure of the change in lateral dimensions on elastically

loading an object and it is defined as  $\mu_{\text{Poisson}} = (1/2) - 3/(6K_{\infty}/G_{\infty} + 2)$ . The claim by Novikov and Sokolov<sup>1</sup> that the fragility of a liquid is determined just by the Poisson ratio of its glass erroneously suggests that controlled modification of the composition, which determines a glass melt's  $m$ , can be used to control its Poisson ratio. ■

**Spyros N. Yannopoulos\*, G. P. Johari†**

\*Foundation for Research and Technology Hellas, Institute of Chemical Engineering and High Temperature Chemical Processes, PO Box 1414, 26504 Patras, Greece

e-mail: sny@iceht.forth.gr

†Department of Materials Science and Engineering, McMaster University, Hamilton, Ontario L8S 4L7, Canada

- Novikov, V. N. & Sokolov, A. P. *Nature* **431**, 961–963 (2004).
- Malek, J. & Shanelova, J. *J. Non-Cryst. Solids* **243**, 116–122 (1999).
- Nemanich, R. J., Gorman, M. & Solin, S. A. *Solid State Commun.* **21**, 277–280 (1977).
- Huang, D. & McKenna, G. B. *J. Chem. Phys.* **114**, 5621–5630 (2001).
- Schroeder, J. *et al. Electron. Lett.* **23**, 1128–1130 (1987).
- Hasz, W. C., Whang, J. H. & Moynihan, C. T. *J. Non-Cryst. Solids* **161**, 127–132 (1993).
- Torell, L. M. *J. Non-Cryst. Solids* **56**, 93–98 (1983).
- Malek, J. *Thermochim. Acta* **311**, 183–198 (1998).
- Nemanich, R. J. *Phys. Rev. B* **16**, 1655–1674 (1977).
- Malek, J. & Shanelova, J. *J. Non-Cryst. Solids* **351**, 3458–3467 (2005).
- Bernatz, K. M., Echeverria, I., Simon, S. L. & Plazek, D. J. *J. Non-Cryst. Solids* **307–310**, 790–801 (2002).
- Ito, Y., Kashida, S. & Murase, K. *Solid State Commun.* **65**, 449–452 (1988).
- Moynihan, C. T., Balitactac, N., Boone, L. & Litovitz, T. A. *J. Chem. Phys.* **55**, 3013–3019 (1971).
- Pohl, R. O., Liu, X. & Thompson, E. *Rev. Mod. Phys.* **74**, 991–1013 (2002).
- Hassan, K., Borjesson, L. & Torell, L. M. *J. Non-Cryst. Solids* **172–174**, 154–160 (1994).
- Macedo, P. B. & Napolitano, A. *J. Chem. Phys.* **49**, 1887–1895 (1968).
- Heiman, D., Hamilton, D. S. & Hellwarth, R. W. *Phys. Rev. B* **19**, 6583–6592 (1979).
- Schilling, F. R., Hausser, M., Sinogeikin, S. V. & Bass, J. D. *Contrib. Mineral. Petrol.* **141**, 297–306 (2001).
- Avramov, I. *J. Non-Cryst. Solids* **238**, 6–10 (1998).
- Lörösch, J., Couzi, L. M., Pelous, J., Vacher, R. & Levasseur, A. *J. Non-Cryst. Solids* **69**, 1–25 (1984).
- El-Moneim, A. *Physica B* **325**, 319–332 (2003).
- Mazurin, O. V., Streltsina, M. V. & Shvaiko-Shvaikovskaya, T. P. *Handbook of Glass Data* (Elsevier, Amsterdam, 1983).
- Komatsu, T. *J. Non-Cryst. Solids* **185**, 199–202 (1995).
- Chen, H. S., Krause, J. T. & Coleman, E. *J. Non-Cryst. Solids* **18**, 157–171 (1975).
- Perera, D. N. *J. Phys. Condens. Matter* **11**, 3807–3812 (1999).
- Wang, W. H. *et al. Appl. Phys. Lett.* **79**, 3947–3949 (2001).
- Zhang, B., Wang, R. J., Zhao, D. Q., Pan, M. X. & Wang, W. H. *Phys. Rev. B* **70**, 224208 (2004).
- Lambson, E. F. *et al. Phys. Rev. B* **33**, 2380–2385 (1986).
- Wang, W. H., Dong, C. & Shek, C. H. *Mater. Sci. Eng. R* **44**, 45–89 (2004).
- Chen, H. S., Krause, J. T. & Coleman, E. *J. Non-Cryst. Solids* **18**, 157–171 (1975).
- Wang, W. H., Lewandowski, J. J. & Greer, A. L. *J. Mater. Res.* **20**, 2037–2313 (2005).
- Tanner, L. E. & Ray, R. *Scripta Metall.* **11**, 783–789 (1977).
- Zhao, Z. F. *et al. Appl. Phys. Lett.* **82**, 4699–4701 (1983).
- Chrysikos, G. D. *J. Non-Cryst. Solids* **172–174**, 378–383 (1994).
- El-Adawy, A. & Moustafa, Y. *J. Phys. D* **32**, 2791–2796 (1999).

doi:10.1038/nature04967

# P2X receptors as cell-surface ATP sensors in health and disease

Baljit S. Khakh<sup>1,2</sup> & R. Alan North<sup>3</sup>

**P2X receptors are membrane ion channels activated by the binding of extracellular adenosine triphosphate (ATP). For years their functional significance was consigned to distant regions of the autonomic nervous system, but recent work indicates several further key roles, such as afferent signalling, chronic pain, and in autocrine loops of endothelial and epithelial cells. P2X receptors have a molecular architecture distinct from other ion channel protein families, and have several unique functional properties.**

The nucleotide ATP is the source of free energy used to maintain order within most cells, and nucleotides are building blocks for key intracellular molecules. Clever experiments some fifty years ago also indicated that some nerves released ATP<sup>1</sup>, leading to the proposal—in 1972—of ‘purinergic nerves’<sup>2</sup>. This was significant because it clearly explicated an extracellular role for an otherwise intracellular molecule. There is now strong molecular, cellular and systems-level evidence for extracellular ATP signalling during physiological processes, placing ATP signalling at centre stage as a widespread facet of cell sensing in diverse organisms. Apparently, evolution has assured that ATP is indispensable inside cells, and as a signalling molecule between them.

Cell-surface nucleotide receptors are called P2 purinoceptors<sup>3</sup>. The letter ‘P’ indicates that they are activated by purines (and in some cases pyrimidines), and the number ‘2’ discriminates them from P1 (now known as A<sub>1</sub>, A<sub>2A</sub>, A<sub>2B</sub> and A<sub>3</sub>) receptors for the nucleoside adenosine. Like other transmitters, ATP activates a family of metabotropic receptors (P2Y) as well as ionotropic receptors (P2X) (Fig. 1a). Metabotropic receptors couple to intracellular second-messenger systems through heteromeric G-proteins. In contrast, P2X receptors contain intrinsic pores that switch conformation from closed to open on binding ATP, allowing ions to flow. The flow of ions is a key step in signalling, because it changes the transmembrane potential as well as local ion concentrations. Additionally, ATP is rapidly degraded in extracellular space by ecto-enzymes<sup>4</sup> to ADP, AMP and adenosine, which also produce physiological responses at G-protein-coupled receptors of the P2Y and adenosine type (Fig. 1a).

The focus of this review is recent discoveries concerning P2X receptors. Two general points that arise from these studies are (1) that P2X receptors are involved in pathophysiological processes throughout the body, and (2) that they differ from other ionotropic receptors in important ways. P2Y receptors are not covered here, and readers are referred to a detailed review of these molecules (ref. 3).

## A new receptor architecture

Early work on ATP-gated channels in nerves and smooth muscles provided evidence for rapid activation kinetics, cation-permeable pores, and significant Ca<sup>2+</sup> permeability<sup>5</sup>. The cloning of genes encoding P2X receptors indicated that the receptors defined a new protein family<sup>6,7</sup> (Figs 1b and 2). P2X genes are found in species from several phyla<sup>5</sup>, but they seem to be absent from the genomes of

*Drosophila* and *Caenorhabditis elegans*. The seven human P2X subunits range from 388 (P2X<sub>4</sub>) to 595 (P2X<sub>7</sub>) amino acids in length (Fig. 2). Each subunit has two hydrophobic, putative membrane-spanning segments (TM1 and TM2) separated by an ectodomain that contains ten conserved cysteine residues that form disulphide bonds<sup>5</sup>. Thus, the amino and carboxy termini are intracellular, and most of the molecule (about 280 amino acids) forms an extracellular loop (Figs 1b and 2); this resembles the overall form of epithelial Na<sup>+</sup> channels (ENaC) and acid-sensing channels (ASIC). Blue-native polyacrylamide gel electrophoresis<sup>8</sup> and atomic force microscopy<sup>9</sup> suggest that P2X channels contain three subunits. Substituted-cysteine accessibility experiments indicate that TM1 and TM2 both may contribute to the pore<sup>5</sup>; the resulting six  $\alpha$ -helices—from three subunits—could form a permeation pathway.

## Biophysical properties of P2X receptors

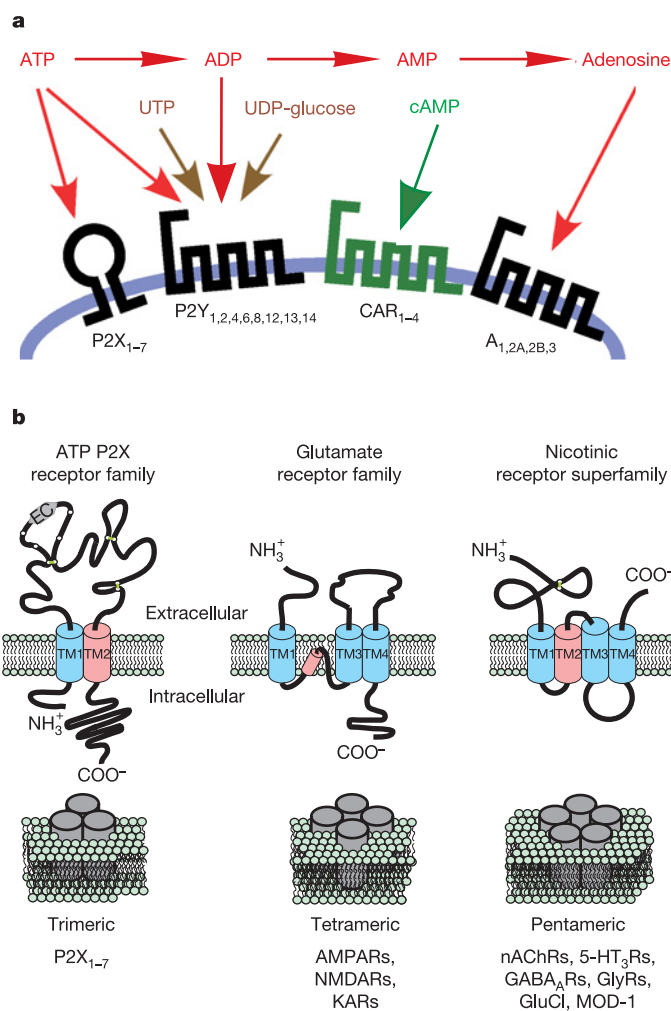
Three molecules of ATP seem to bind to the extracellular portions of P2X receptors<sup>5,10</sup>. Mutagenesis experiments indicate that conserved lysines near the extracellular ends of TM1 and TM2 may contribute to ATP binding<sup>5</sup>, and that aromatic residues within the extracellular loop may coordinate the adenine group of ATP<sup>11</sup>; more direct structural studies are keenly awaited. Electrophysiological experiments on P2X receptors also reveal functionally important interfaces between neighbouring subunits<sup>10</sup>. Moreover, Zn<sup>2+</sup> binds allosterically to histidine residues at subunit interfaces to modulate ATP-evoked channel opening<sup>12</sup>. Studies on nicotinic receptors indicate that two acetylcholine (ACh) molecules bind via electrostatic interactions within aromatic binding pockets at subunit interfaces: aspects of the binding site may undergo a conformational change to “close around the bound ACh molecule”<sup>13</sup>. In the case of ionotropic glutamate receptors, the transmitter binds in a clam shell formed between the S1 and S2 domains of individual subunits in the tetramer, and the clam shell closes around the bound glutamate<sup>14</sup>.

One might expect that a narrowing pore is formed by a common subunit interface located centrally within P2X receptors, as is the case for the nicotinic superfamily and ionotropic glutamate receptors. Most P2X channels are cation-selective and—like members of the nicotinic superfamily and ionotropic glutamate receptors—they discriminate only poorly among different cations. P2X<sub>5</sub> receptors are significantly chloride permeable<sup>15</sup>, recalling  $\gamma$ -aminobutyric acid receptor A (GABA<sub>A</sub>), glycine, glutamate-gated chloride (GluCl) and

<sup>1</sup>Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK. <sup>2</sup>Departments of Physiology and Neurobiology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California 90095-1751, USA. <sup>3</sup>Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK.



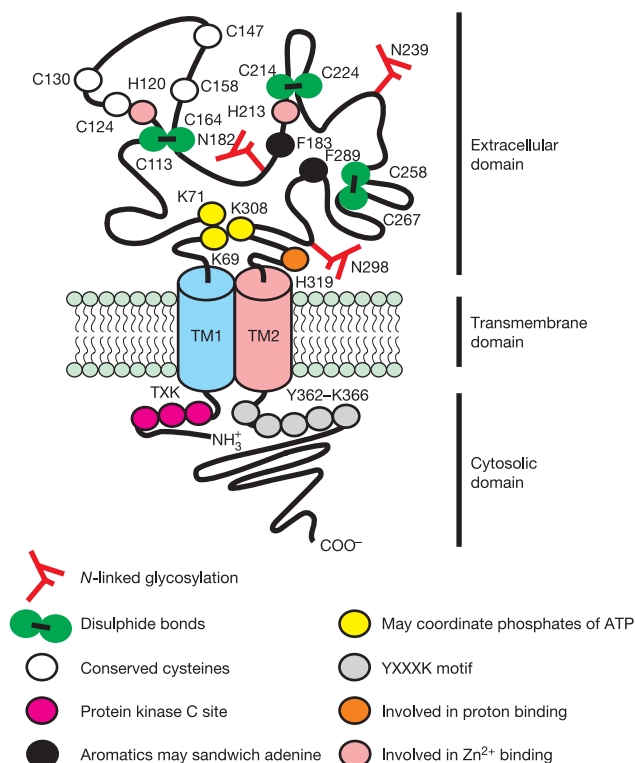
MOD-1 receptors from the nicotinic superfamily. Single-channel recordings suggest that a low field-strength, cation-binding site exists in the P2X<sub>2</sub> pore<sup>16</sup>. Some P2X receptors also show significant permeability to divalent cations, with Ca<sup>2+</sup> fluxes greater than most nicotinic family and glutamate-gated receptors, and similar to those found in *N*-methyl-D-aspartate (NMDA) receptors<sup>17</sup>. These data indicate a selectivity filter in P2X channels that favours the flow of selected ions over others. Polar threonine and serine residues group on one face of  $\alpha$ -helical TM2, and their side chains from three subunits projecting into the pore may act as surrogate water molecules<sup>17</sup>. Thus, the P2X<sub>2</sub> selectivity filter may consist of, at least in part, a small polar stretch within TM2, extracellular to the gate.



**Figure 1 | ATP signalling and families of mammalian ionotropic receptors.** **a**, Receptors for extracellular nucleotides. ATP is converted to adenosine 5'-diphosphate (ADP), adenosine 5'-monophosphate (AMP), and adenosine by ecto-enzymes of several classes<sup>4</sup>. These are the nucleotide triphosphate diphosphohydrolases (NTPDases), ecto-nucleotide pyrophosphatase/phosphodiesterases (E-NPP), ecto-alkaline phosphatases, and ecto-5'-nucleotidases. Nucleotide/nucleoside receptors in vertebrates (black) include P2X, P2Y and adenosine receptors. ATP acts on P2X and some P2Y receptors (G-protein-coupled, seven transmembrane domain); some P2Y receptors are activated best by ADP, uridine 5'-triphosphate (UTP), and uridine 5'-diphosphate (UDP)-glucose. Adenosine acts at four adenosine receptors. In *Dictyostelium* (green), cyclic adenosine 5'-monophosphate (cAMP) acts at a family of four cAMP receptors (CAR). **b**, Cartoon representations of subunit arrangement and proposed topology for mammalian ionotropic receptors. EC, extracellular domain; TM1–TM4, transmembrane domains 1–4; KAR, kainate receptor.

The narrowest part of the pore that moves to allow ion flow—the activation gate—in P2X channels may be near a conserved glycine residue in TM2 (Gly 342 in rat P2X<sub>2</sub>), but decisive evidence is awaited. Glycine, which lacks a side chain, may favour local conformational flexibility. For example, in certain K<sup>+</sup> channels glycine forms a gating hinge<sup>18</sup>, and in nicotinic channels glycines are found at the intra- and extra-cellular ends of TM2, where they may permit flexibility<sup>19</sup>. The conserved glycine in P2X TM2 is always followed by a hydrophobic residue, with a second hydrophobic residue positioned one helical turn away<sup>5</sup>. Might the activation gate be formed here? This would be consistent with the finding that mutation of the conserved glycine locks the channel into distinct permeability states<sup>20</sup>. Moreover, hydrophobic residues seem to be a common feature for the gates of channels with known structures. For example, the gates of nicotinic, mechanosensitive and bacterial inward-rectifier K<sup>+</sup> channels are all formed by hydrophobic residues<sup>18</sup>. By restricting the flow of water, hydrophobic residues may impede ion movement past the gate<sup>21</sup>. In P2X receptors, cysteine-substitution experiments indicate that TM1 and TM2 line the pore<sup>5</sup>, but so far they fall short of pinpointing the activation gate.

Once past the gate, permeant ions enter the cytosol. Fluorescence resonance energy transfer (FRET) data indicate that the cytosolic domains of P2X<sub>2</sub> receptors may narrow<sup>22</sup> from their origins at the plasma membrane to their tips by ~10 Å. This implies the existence of an ordered cytosolic domain in P2X channels, which may contain additional barriers/exits for ions, as is the case for cytosolic domains of nicotinic superfamily receptors<sup>23</sup>. The sequences of the C termini of P2X receptors vary widely among the seven subunits<sup>5</sup>, although a conserved YXXXK sequence (where "X" is any amino acid) in the juxtamembrane region<sup>5,24</sup> is involved in membrane retention (Fig. 2). Other motifs are probably involved in endocytosis<sup>25</sup>, permeability changes<sup>26</sup>, binding of lipopolysaccharides<sup>27</sup>, and interactions with other proteins<sup>28,29</sup>.



**Figure 2 | Summary of P2X receptor structure-function studies.** Topology and key features of P2X receptor subunits. This hypothetical view is based mainly on experiments with P2X<sub>1</sub> and P2X<sub>2</sub> receptors<sup>5</sup>. The numbers are for rat P2X<sub>2</sub> receptors.

## Central nervous system P2X receptors and neuropathic pain

P2X receptor subunits are widely expressed in the nervous system—both in neurons and glia<sup>5,30</sup>—with P2X<sub>2</sub>, P2X<sub>4</sub> and P2X<sub>6</sub> being the most abundant in neurons. At some central synapses, ATP is a fast neurotransmitter<sup>5,31</sup> where it may be released with other transmitters<sup>32,33</sup>, as is the case in the periphery<sup>5</sup>. In general, recalling findings for fast nicotinic and 5-hydroxytryptamine (5-HT) brain synapses, ATP synaptic currents in the brain are small, and only occur in a subpopulation of neurons. The functional significance at central synapses might be related to the high Ca<sup>2+</sup> fluxes allowed by P2X receptors, even at resting membrane potentials: Ca<sup>2+</sup> entry during ATP synaptic transmission affects the frequency dependence for the induction of synaptic plasticity<sup>34</sup>. Whereas gene knockout experiments have confirmed a role of specific subunits at a neuroeffector junction (P2X<sub>1</sub>; ref. 35) and in myenteric synapses (P2X<sub>2</sub>; ref. 36), the molecular composition of synaptically activated P2X receptors in brain neurons remains poorly defined.

Antibody labelling has revealed that P2X<sub>2</sub> and P2X<sub>4</sub> subunits are found in excitatory synapses in the brain<sup>30</sup>. Unlike  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) receptor subunits, P2X<sub>2</sub> and P2X<sub>4</sub> subunits are found at the periphery of the postsynaptic density; that is, in regions where AMPA receptor subunits are of low density<sup>29,30</sup>. This might indicate that synaptic P2X receptors are only activated during bouts of action potential firing, perhaps when synaptically released ATP can spill out to activate P2X channels. We presume that the location of P2X receptors to the periphery of postsynaptic densities arises from specific protein interactions: one such protein that colocalizes and interacts with P2X<sub>2</sub> receptors in postsynaptic membranes is Fe65 (ref. 29), better known as a partner to  $\beta$ -amyloid precursor protein. Perhaps the number of synaptic P2X channels depends critically on the rates of insertion and removal by regulated trafficking.

In addition to mediating postsynaptic responses, there is now overwhelming evidence for presynaptic P2X receptors throughout the central nervous system<sup>5,37,38</sup>. In general, an increase in transmitter release probability is observed. The responses are triggered by calcium ions, which enter through the P2X receptor pores themselves

or through voltage-dependent Ca<sup>2+</sup> channels, which are activated as a consequence of the P2X-receptor-mediated nerve terminal depolarization. Presynaptic P2X channels may be activated physiologically in some synapses<sup>38–40</sup>. In the nucleus of the solitary tract (NTS), activation of presynaptic P2X channels evoked large-amplitude miniature synaptic currents<sup>41</sup>. This indicates that Ca<sup>2+</sup> entry through presynaptic P2X channels may trigger multivesicular release<sup>41</sup>, or release of glutamate from an otherwise ‘silent’ set of vesicles or terminals.

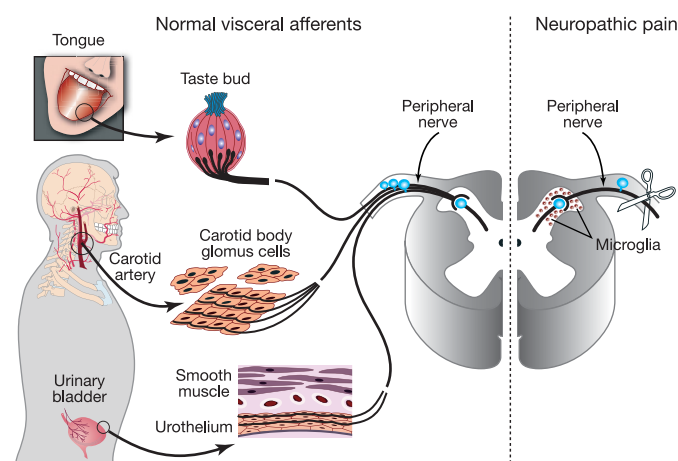
Presynaptic actions form one component of ATP effects in the hippocampus, specifically in the feed-forward circuit formed between CA3 pyramidal neurons, GABAergic interneurons, and output CA1 pyramidal neurons. Thus, presynaptic P2X<sub>2</sub> receptors increase glutamate release onto interneurons<sup>40</sup>, whereas postsynaptic P2Y<sub>1</sub> receptors depolarize interneurons<sup>42,43</sup> and astrocyte P2Y receptors evoke glutamate release onto interneurons<sup>42</sup>. However, because there are few pre- or post-synaptic ATP receptors on CA1 pyramidal neurons, the net effect on output neurons is dominated by heightened GABAergic synaptic inhibition. Thus, in the stratum radiatum region of the hippocampus, although the cellular effects of ATP are excitatory, the overall result on the network is dominated by increased inhibition. ATP may act as a ‘physiological brake’ to runaway excitation within this feed-forward circuit<sup>42</sup>.

Substantial ATP is released during repetitive firing of neurons<sup>44</sup> and from glia<sup>45</sup>, as well as during injury<sup>46,47</sup> and ischaemia<sup>48</sup> of the brain, and the integrative actions of ATP released from these multiple sources are beginning to be understood. P2X<sub>4</sub> subunits have recently been implicated in pain sensation, but, in this case, it seems that the critical sites of expression are brain microglia rather than neurons. The tactile allodynia (whereby hitherto innocuous stimuli to the skin now elicit intense pain) that develops several days after ligation of a spinal nerve is much reduced by blocking the expression of P2X<sub>4</sub> subunits in the dorsal horn using intrathecal antisense oligonucleotides<sup>49</sup>. This also blocked the activation of dorsal horn microglia. Remarkably, intrathecal administration of a solution containing cultured brain microglia also produced the allodynia, but only if they had been pretreated with ATP<sup>49</sup>. It is suggested that ATP—by acting on P2X<sub>4</sub> receptors—causes microglia to release brain-derived neurotrophic factor (BDNF)<sup>50</sup>. This acts on nearby neurons to alter the expression of an anion transporter; the altered chloride gradient causes the key transmitter GABA to switch its effect from inhibition to excitation. If it occurred in the appropriate neurons, this could underlie the increased sensitivity that is a feature of allodynia. Together, these studies highlight a new systems-level mechanism involving endogenously released ATP acting on P2X<sub>4</sub> receptors, as well as a new target to treat neuropathic pain.

## Physiological and pathological sensations

ATP elicits acute pain when applied to human skin<sup>51,52</sup>, and the cell bodies of sensory neurons were among the first sites where it was shown to activate ligand-gated channels<sup>7</sup>. Thus, a role for ATP in pain sensing was buttressed by the discovery that the P2X<sub>3</sub> subunit was expressed almost exclusively in a subset of primary afferents implicated in nociception<sup>53,54</sup>, and that a predominant form of receptor in such cells is a heteromeric channel comprising P2X<sub>2</sub> and P2X<sub>3</sub> subunits<sup>54,55</sup>. But the behavioural response to acute painful stimuli (mechanical stimulation and heat) is unaffected in the P2X<sub>3</sub> knockout mice, arguing against ATP acting simply on P2X<sub>3</sub> as an acute pain-evoking substance<sup>56,57</sup>. More intriguing is the accumulating evidence that P2X<sub>3</sub> receptors are involved in the chronic or persistent nociceptive behaviour that follows nerve injury or inflammation, and may have a more widespread role in normal visceral afferent sensation.

Mechanical allodynia can be readily measured in experimental animals, and is widely used as a model of the chronic neuropathic pain in man that occurs in, for example, diabetes or AIDS. Mechanical allodynia is greatly reduced in mice with deleted P2X<sub>3</sub> receptor genes<sup>56,57</sup>, in rats that have been treated with intrathecal antisense



**Figure 3 | Sensory transduction by P2X receptors.** Left: Schematic illustration to show sensory nerve pathways from tongue taste buds, the carotid body and the bladder wall. In each case, evidence implicates ATP as the transmitter released by the sensory cell to initiate an action potential in the sensory nerve. Right: In response to injury to a peripheral nerve, microglia become activated in the dorsal horn of the spinal cord. It is proposed that ATP acts on P2X<sub>4</sub> receptors on these microglia to elicit the release of BDNF, which then acts on neighbouring neurons to change their response (and thus the perception) of incoming sensory signals. Further details are provided in the text.



oligonucleotides that reduce expression of P2X<sub>3</sub> receptors<sup>58</sup>, and in rats receiving pharmacological antagonists selective for P2X receptors that contain P2X<sub>3</sub> subunits<sup>58,59</sup>. Such antagonists are effective also in other models of neuropathic and inflammatory pain, and several pharmaceutical companies are now pursuing these actions with a view to novel therapeutics. P2X<sub>3</sub>-receptor knockout mice have further defects in afferent sensation (Fig. 3). Those lacking P2X<sub>3</sub> subunits have impaired ability to sense bladder filling (at least when anaesthetized)<sup>56</sup>. Together with earlier work showing that bladder distension releases ATP in the urothelium (the layer of the bladder wall that contains the peripheral endings of sensory nerves), this suggests that ATP forms the link between sensory stimulus (in this case mechanical) and sensory nerve activation (Fig. 3). ATP, and analogues that activate P2X<sub>3</sub>-subunit-containing receptors, also excite afferent nerves in the wall of the intestine<sup>60</sup>; if ATP were to have a similar role in distended and inflamed intestinal mucosa, then blockade of P2X<sub>3</sub> receptors may be a valuable therapeutic approach in irritable bowel syndrome<sup>61</sup>.

A similar link has recently been adduced for taste receptors in the tongue (Fig. 3), where the taste buds release ATP that then activates P2X<sub>2/3</sub> heteromeric receptors on gustatory nerves<sup>62</sup>. The carotid body is a collection of specialized epithelial cells in contact with arterial blood found near the bifurcation of the carotid artery (Fig. 3). These cells sense the blood oxygen and carbon dioxide concentrations. The response to hypoxia (though not hypercapnia) involves ATP release from sensory glomus cells onto terminals of the carotid sinus nerve<sup>63</sup>; although these sensory nerves express both P2X<sub>2</sub> and P2X<sub>3</sub> subunits, in this case studies on knockout mice indicate that the P2X<sub>2</sub> subunit has the predominant role<sup>63</sup>. ATP release is also involved in central chemosensory transduction<sup>64</sup>, and other signalling roles for P2X receptors need to be tested.

### P2X receptors in cell permeabilization and inflammation

One of the more remarkable actions of extracellular ATP is its ability to 'permeabilize' certain cells. This was first observed in peritoneal mast cells<sup>65</sup>; it is now more commonly studied in macrophages and related cells as the ATP-induced uptake of dyes (such as ethidium and YO-PRO-1)<sup>5</sup>, which become fluorescent when they bind to intracellular nucleic acids. An important role in inflammation was inferred from the finding that ATP also elicited the processing and release of pro-inflammatory cytokines (particularly interleukin-1 and interleukin-6) from lipopolysaccharide-primed macrophages by acting on a receptor with similar properties<sup>5</sup>. This was initially called the P2Z receptor<sup>66</sup>; its distinct properties were (1) low sensitivity to ATP relative to other P2X receptors (typically 100  $\mu$ M to 3 mM needed), (2) high sensitivity to the analogue 2',3'-O-(4-benzoylbenzoyl)ATP (BzATP) relative to ATP, and (3) marked potentiation of the response by reducing the concentration of extracellular divalent cations. At the time of gene cloning it was recognized that these properties are closely shared by the homomeric P2X<sub>7</sub> receptor<sup>67</sup>, which also has a widespread distribution in macrophages and other cells of the immune system. Macrophages from P2X<sub>7</sub> gene-deficient mice do not release interleukin-1 $\beta$  in response to ATP<sup>68</sup>. Such mice develop much reduced arthritis on induction by collagen injection into their joints<sup>69</sup>, and show deficient periosteal bone formation due to absence of the receptor in osteoclasts (which behave as bone macrophages)<sup>70</sup>.

The relationship remains enigmatic between the 'permeabilization' or 'pore formation' and the downstream actions of P2X<sub>7</sub> receptors such as interleukin release. The opening of the cation channel of the P2X<sub>7</sub> receptor occurs within milliseconds of applying the agonist ATP, and this is broadly similar to the other P2X receptors. When the agonist application is continued for several seconds, some P2X receptors<sup>20,71</sup>, but notably P2X<sub>7</sub> receptors<sup>67</sup>, become permeable to the large cation *N*-methyl-D-glucamine, as determined by direct biophysical measures of ion reversal potential. This time course is similar to that observed for dye uptake, but recent evidence suggests

that the two phenomena may not be mechanistically related for P2X<sub>7</sub> receptors<sup>72</sup>. The dye uptake may reflect signal transduction from P2X<sub>7</sub> receptors to distinct membrane proteins; for instance, inhibition of a P-glycoprotein has been suggested<sup>73</sup>. Moreover, there is no evidence that the opening of the intrinsic cation channel, or the pathway through which dyes enter cells, are necessary for the release of interleukins evoked by ATP. It seems likely that these aspects of P2X<sub>7</sub> signalling pass through related proteins of the complex in which the receptor<sup>28</sup> and downstream effectors reside.

P2X<sub>7</sub>-receptor-deficient mice also show reduced or absent behavioural responses to inflammation of the paw, as well as tactile allodynia<sup>74</sup>. A component of these effects seems to result from a reduced release of pro-inflammatory interleukin-1 $\beta$  in the paw itself. However, brain microglia also express abundant P2X<sub>7</sub> receptors when activated by an injurious stimulus, and a central component at some point in the nociceptive pathway also seems likely to contribute.

### Towards broader physiological roles

P2X receptors are widely expressed, not only within nervous tissue. Progress on several fronts has revealed the diversity of cellular signalling. P2X<sub>4</sub> subunits, for example, are found in very many tissues, and their role in epithelial and endothelial tissues has attracted particular interest. In epithelia, they are abundant in salivary glands and bronchiolar epithelium. In the bronchioles, they seem to have a key role in maintaining the beating of cilia that moves the mucus layer, and thus helps to clear the airways of pathogens; the receptor has functional features suggesting P2X<sub>4</sub> and/or P2X<sub>7</sub> subunits<sup>75</sup>. It has been suggested that ATP forms part of an autocrine loop, and that the calcium entry through activated P2X receptors would be of therapeutic benefit in cystic fibrosis.

P2X<sub>1</sub> receptors are abundantly expressed in blood vessel smooth muscle and platelets, and a role in the regulation of blood pressure or haemostasis was postulated. However, knockout mice showed little deficit in blood pressure or clotting; the P2X<sub>1</sub> knockout male mice were infertile, owing to a deficit in P2X<sub>1</sub>-receptor-mediated purinergic transmission in the vas deferens<sup>35</sup>. The first described phenotype of the P2X<sub>4</sub> knockout mice was an increase in blood pressure. In the endothelium of the vasculature, the rise in intracellular calcium resulting from shear stress is due to calcium entry through P2X<sub>4</sub> receptors<sup>76</sup>. In the absence of P2X<sub>4</sub>, less nitric oxide is released and the normal vasodilatation evoked by acute increases in blood flow is blunted. The knockout mice have a higher blood pressure than their wild-type controls, and do not show adaptive vascular remodelling in response to a chronic decrease in blood flow<sup>76</sup>. The discovery of antagonist molecules effective at P2X<sub>4</sub> receptors would greatly facilitate the further elucidation of their physiological roles, and may lead to novel therapeutics. Bone metabolism is impaired in mice with a disrupted P2X<sub>7</sub> receptor gene<sup>70</sup>. The receptor is normally expressed on both osteoblasts and osteoclasts: its absence results in a unique skeletal phenotype, with excessive resorption of trabecular bone. Thus, the P2X<sub>7</sub> receptor is being pursued as a therapeutic target in osteoporosis.

### Outlook

The widespread expression of P2X receptors in cells prepares us for the possibility that ATP may contribute generally to physiological processes in phylogenetically diverse organisms, raising several classes of question.

In the first class, what is the true extent of the family? For example, why do the genomes of slime mould and the trematode *Schistosoma mansoni* contain P2X genes<sup>77</sup>, whereas those of the nematode *C. elegans* and the fruitfly do not? Might these organisms utilize a presently unknown family of ionotropic ATP receptors? Can the P2X channel in *S. mansoni* be targeted to treat schistosomiasis? In view of the key signalling role of extracellular cyclic AMP, what physiological functions

could P2X channels serve in slime mould? Will bioinformatics yet identify other organisms that contain P2X channels?

Questions of the second class deal with the molecular *modus operandi*. How do P2X receptors couple the energy of ATP binding to channel opening? How does the selectivity filter in P2X channels work, and select for  $\text{Ca}^{2+}$  over the more abundant  $\text{Na}^{+}$ ? What is the atomic-scale structure of a P2X receptor? Are there general principles of permeation and gating that are conserved between structurally distinct ionotropic receptors? What are the molecular mechanisms of subunit-specific channel rectification and desensitization? Can the intrinsic millisecond time-scale function of P2X channels be exploited to generate new genetically encoded reporters of ATP release? Can engineered P2X channels with optimized sensitivity, kinetics and permeability be used in biosensor applications, such as remote sensing? Can the permeabilizing activity of P2X<sub>7</sub> receptors be exploited in gene therapy approaches to treat tumours?

The progressive discoveries of novel phenotypes in P2X-receptor knockout mice are driving a third class of question. How widespread is ATP as a transmitter activating primary afferent nerves? Does this reveal some biological principles in sensory processing? Is ATP released along with every other transmitter from vesicles? The study of P2X receptors has been seriously handicapped by a lack of pharmacological tools. New drugs that selectively block P2X receptors are now becoming available. Their application promises to lead us into areas of physiology hitherto not considered and, hopefully, into novel therapeutics for human disease.

- Holton, P. The liberation of adenosine triphosphate on antidromic stimulation of sensory nerves. *J. Physiol. (Lond.)* **145**, 494–504 (1959).
- Burnstock, G. Purinergic nerves. *Pharmacol. Rev.* **24**, 509–581 (1972).
- Ralevic, V. & Burnstock, G. Receptors for purines and pyrimidines. *Pharmacol. Rev.* **50**, 413–492 (1998).
- Zimmermann, H. Extracellular metabolism of ATP and other nucleotides. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **362**, 299–309 (2000).
- North, R. A. Molecular physiology of P2X receptors. *Physiol. Rev.* **82**, 1013–1067 (2002).
- Valera, S. *et al.* A new class of ligand-gated ion channel defined by P2X receptor for extracellular ATP. *Nature* **371**, 516–519 (1994).
- Brake, A. J., Wagenbach, M. J. & Julius, D. New structural motif for ligand-gated ion channels defined by an ionotropic ATP receptor. *Nature* **371**, 519–523 (1994).
- Nicke, A. *et al.* P2X<sub>1</sub> and P2X<sub>3</sub> receptors form stable trimers: a novel structural motif of ligand-gated ion channels. *EMBO J.* **17**, 3016–3028 (1998).
- Barrera, N. P., Ormond, S. J., Henderson, R. M., Murrell-Lagnado, R. D. & Edwardson, J. M. Atomic force microscopy imaging demonstrates that P2X<sub>2</sub> receptors are trimers but that P2X<sub>6</sub> receptor subunits do not oligomerize. *J. Biol. Chem.* **280**, 10759–10765 (2005).
- Jiang, L. H. *et al.* Subunit arrangement in P2X receptors. *J. Neurosci.* **23**, 8903–8910 (2003).
- Roberts, J. A. & Evans, R. J. ATP binding at human P2X<sub>1</sub> receptors. Contribution of aromatic and basic amino acids revealed using mutagenesis and partial agonists. *J. Biol. Chem.* **279**, 9043–9055 (2004).
- Nagaya, N., Tittle, R. K., Saar, N., Dellal, S. S. & Hume, R. I. An intersubunit zinc binding site in rat P2X<sub>2</sub> receptors. *J. Biol. Chem.* **280**, 25982–25993 (2005).
- Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346**, 967–989 (2005).
- Madden, D. R. The structure and function of glutamate receptor ion channels. *Nature Rev. Neurosci.* **3**, 91–101 (2002).
- Bo, X. *et al.* Pharmacological and biophysical properties of the human P2X<sub>5</sub> receptor. *Mol. Pharmacol.* **63**, 1407–1416 (2003).
- Ding, S. & Sachs, F. Ion permeation and block of P2X<sub>2</sub> purinoceptors: single channel recordings. *J. Membr. Biol.* **172**, 215–223 (1999).
- Egan, T. M. & Khakh, B. S. Contribution of calcium ions to P2X channel responses. *J. Neurosci.* **24**, 3413–3420 (2004).
- Doyle, D. A. Structural changes during ion channel gating. *Trends Neurosci.* **27**, 298–302 (2004).
- Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **424**, 949–955 (2003).
- Khakh, B. S., Bao, X., Labarca, C. & Lester, H. A. Neuronal P2X receptor-transmitter-gated channel channels change their ion selectivity in seconds. *Nature Neurosci.* **2**, 322–330 (1999).
- Doyle, D. A. Molecular insights into ion channel function. *Mol. Membr. Biol.* **21**, 221–225 (2004).
- Fisher, J. A., Girdler, G. & Khakh, B. S. Time resolved measurement of state specific P2X ion channel cytosolic gating motions. *J. Neurosci.* **24**, 10475–10487 (2004).
- Miyazawa, A., Fujiyoshi, Y., Stowell, M. & Unwin, N. Nicotinic acetylcholine receptor at 4.6 Å resolution: transverse tunnels in the channel wall. *J. Mol. Biol.* **288**, 765–786 (1999).
- Chaumont, S., Jiang, L. H., Penna, A., North, R. A. & Rassendren, F. Identification of a trafficking motif involved in the stabilization and polarization of P2X receptors. *J. Biol. Chem.* **279**, 29628–29638 (2004).
- Royle, S. J., Bobanovic, L. K. & Murrell-Lagnado, R. D. Identification of a non-canonical tyrosine-based endocytic motif in an ionotropic receptor. *J. Biol. Chem.* **277**, 35378–35385 (2002).
- Eickhorst, A., Berson, A., Cockayne, D., Lester, H. A. & Khakh, B. S. Control of P2X<sub>2</sub> channel permeability by the cytosolic domain. *J. Gen. Physiol.* **120**, 119–131 (2002).
- Denlinger, L. C. *et al.* Cutting edge: the nucleotide receptor P2X<sub>7</sub> contains multiple protein- and lipid-interaction motifs including a potential binding site for bacterial lipopolysaccharide. *J. Immunol.* **167**, 1871–1876 (2001).
- Kim, M., Jiang, L. H., Wilson, H. L., North, R. A. & Surprenant, A. Proteomic and functional evidence for a P2X<sub>7</sub> receptor signalling complex. *EMBO J.* **20**, 6347–6358 (2001).
- Masin, M., Kerschensetter, D., Dumke, K., Rubio, M. E. & Soto, F. Fe65 interacts with P2X<sub>2</sub> subunits at excitatory synapses and modulates receptor function. *J. Biol. Chem.* **281**, 4100–4108 (2006).
- Rubio, M. E. & Soto, F. Distinct localisation of P2X receptors at excitatory postsynaptic specializations. *J. Neurosci.* **21**, 641–653 (2001).
- Edwards, F. A., Gibb, A. J. & Colquhoun, D. ATP receptor-mediated synaptic currents in the central nervous system. *Nature* **359**, 144–147 (1992).
- Jo, Y. H. & Role, L. W. Coordinate release of ATP and GABA at *in vitro* synapses of lateral hypothalamic neurons. *J. Neurosci.* **22**, 4794–4804 (2002).
- Jo, Y.-H. & Schlachter, R. Synaptic corelease of ATP and GABA in cultured spinal neurons. *Nature Neurosci.* **2**, 241–245 (1999).
- Pankratov, Y. V., Lalo, U. V. & Krishtal, O. A. Role for P2X receptors in long-term potentiation. *J. Neurosci.* **22**, 8363–8369 (2002).
- Mulryan, K. *et al.* Reduced vas deferens contraction and male infertility in mice lacking P2X<sub>1</sub> receptors. *Nature* **403**, 86–89 (2000).
- Ren, J. *et al.* P2X<sub>2</sub> subunits contribute to fast synaptic excitation in myenteric neurons of the mouse small intestine. *J. Physiol. (Lond.)* **552**, 809–821 (2003).
- Gu, J. G. & MacDermott, A. B. Activation of ATP P2X receptors elicits glutamate release from sensory neuron synapses. *Nature* **389**, 749–753 (1997).
- Khakh, B. S. & Henderson, G. ATP receptor-mediated enhancement of fast excitatory neurotransmitter release in the brain. *Mol. Pharmacol.* **54**, 372–378 (1998).
- Nakatsuka, T. & Gu, J. G. ATP P2X receptor-mediated enhancement of glutamate release and evoked EPSCs in dorsal horn neurons of the rat spinal cord. *J. Neurosci.* **21**, 6522–6531 (2001).
- Khakh, B. S., Gittermann, D., Cockayne, D. A. & Jones, A. ATP modulation of excitatory synapses onto interneurons. *J. Neurosci.* **23**, 7426–7437 (2003).
- Shigetomi, E. & Kato, F. Action potential-independent release of glutamate by  $\text{Ca}^{2+}$  entry through presynaptic P2X receptors elicits postsynaptic firing in the brainstem autonomic network. *J. Neurosci.* **24**, 3125–3135 (2004).
- Bowser, D. N. & Khakh, B. S. ATP excites interneurons and astrocytes to increase synaptic inhibition in neuronal networks. *J. Neurosci.* **24**, 8606–8620 (2004).
- Kawamura, M., Gachet, C., Inoue, K. & Kato, F. Direct excitation of inhibitory interneurons by extracellular ATP mediated by P2Y<sub>1</sub> receptors in the hippocampal slice. *J. Neurosci.* **24**, 10835–10845 (2004).
- Wieraszko, A., Goldsmith, G. & Seyfried, T. N. Stimulation-dependent release of adenosine triphosphate from hippocampal slices. *Brain Res.* **485**, 244–250 (1989).
- Gordon, G. R. *et al.* Norepinephrine triggers release of glial ATP to increase postsynaptic efficacy. *Nature Neurosci.* **8**, 1078–1086 (2005).
- Davalos, D. *et al.* ATP mediates rapid microglial response to local brain injury *in vivo*. *Nature Neurosci.* **8**, 752–758 (2005).
- Wang, X. *et al.* P2X<sub>7</sub> receptor inhibition improves recovery after spinal cord injury. *Nature Med.* **10**, 821–827 (2004).
- Dulla, C. G. *et al.* Adenosine and ATP link  $\text{P}_{\text{CO}_2}$  to cortical excitability via pH. *Neuron* **48**, 1011–1023 (2005).
- Tsuda, M. *et al.* P2X<sub>4</sub> receptors induced in spinal microglia gate tactile allodynia after nerve injury. *Nature* **424**, 778–783 (2003).
- Coull, J. A. *et al.* BDNF from microglia causes the shift in neuronal anion gradient underlying neuropathic pain. *Nature* **438**, 1017–1021 (2005).
- Bleehen, T. & Keele, C. A. Observations on the algogenic actions of adenosine compounds on the human blister base preparation. *Pain* **3**, 367–377 (1977).
- Hamilton, S. G., Warburton, J., Bhattacharjee, A., Ward, J. & McMahon, S. B. ATP in human skin elicits a dose-related pain response which is potentiated under conditions of hyperalgesia. *Brain* **123**, 1238–1246 (2000).
- Chen, C. C. *et al.* A P2X purinoceptor expressed by a subset of sensory neurons. *Nature* **377**, 428–431 (1995).
- Lewis, C. *et al.* Coexpression of P2X<sub>2</sub> and P2X<sub>3</sub> receptor subunits can account for ATP-gated currents in sensory neurons. *Nature* **377**, 432–435 (1995).
- Khakh, B. S., Humphrey, P. P. & Surprenant, A. Electrophysiological properties of P2X-purinoceptors in rat superior cervical, nodose and guinea-pig coeliac neurones. *J. Physiol. (Lond.)* **484**, 385–395 (1995).



56. Cockayne, D. A. *et al.* Urinary bladder hyporeflexia and reduced pain-related behaviour in P2X<sub>3</sub>-deficient mice. *Nature* **407**, 1011–1015 (2000).
57. Souslova, V. *et al.* Warm-coding deficits and aberrant inflammatory pain in mice lacking P2X<sub>3</sub> receptors. *Nature* **407**, 1015–1017 (2000).
58. North, R. A. The P2X<sub>3</sub> subunit: a molecular target in pain therapeutics. *Curr. Opin. Investig. Drugs* **4**, 833–840 (2003).
59. Jarvis, M. F. *et al.* A-317491, a novel potent and selective non-nucleotide antagonist of P2X<sub>3</sub> and P2X<sub>2/3</sub> receptors, reduces chronic inflammatory and neuropathic pain in the rat. *Proc. Natl Acad. Sci. USA* **99**, 17179–17184 (2002).
60. Kirkup, A. J., Booth, C. E., Chessell, I. P., Humphrey, P. P. & Grundy, D. Excitatory effect of P2X receptor activation on mesenteric afferent nerves in the anaesthetised rat. *J. Physiol. (Lond.)* **520**, 551–563 (1999).
61. Galligan, J. J. Enteric P2X receptors as potential targets for drug treatment of the irritable bowel syndrome. *Br. J. Pharmacol.* **141**, 1294–1302 (2004).
62. Finger, T. E. *et al.* ATP signaling is crucial for communication from taste buds to gustatory nerves. *Science* **310**, 1495–1499 (2005).
63. Gourine, A. V. On the peripheral and central chemoreception and control of breathing: an emerging role of ATP. *J. Physiol. (Lond.)* **568**, 715–724 (2005).
64. Gourine, A. V., Llaudet, E., Dale, N. & Spyer, K. M. ATP is a mediator of chemosensory transduction in the central nervous system. *Nature* **436**, 108–111 (2005).
65. Cockcroft, S. & Gomperts, B. D. ATP induces nucleotide permeability in rat mast cells. *Nature* **279**, 541–542 (1979).
66. Gordon, J. L. Extracellular ATP: effects, sources and fate. *Biochem. J.* **233**, 309–319 (1986).
67. Surprenant, A., Rassendren, F., Kawashima, E., North, R. A. & Buell, G. The cytolytic P2Z receptor for extracellular ATP identified as a P2X receptor (P2X<sub>7</sub>). *Science* **272**, 735–738 (1996).
68. Solle, M. *et al.* Altered cytokine production in mice lacking P2X<sub>7</sub> receptors. *J. Biol. Chem.* **276**, 125–132 (2001).
69. Labasi, J. M. *et al.* Absence of the P2X<sub>7</sub> receptor alters leukocyte function and attenuates an inflammatory response. *J. Immunol.* **168**, 6436–6445 (2002).
70. Ke, H. Z. *et al.* Deletion of the P2X<sub>7</sub> nucleotide receptor reveals its regulatory roles in bone formation and resorption. *Mol. Endocrinol.* **17**, 1356–1367 (2003).
71. Virginio, C., MacKenzie, A., Rassendren, F. A., North, R. A. & Surprenant, A. Pore dilation of neuronal P2X receptor channels. *Nature Neurosci.* **2**, 315–321 (1999).
72. Jiang, L. H. *et al.* N-methyl-D-glucamine and propidium dyes utilize different permeation pathways at rat P2X<sub>7</sub> receptors. *Am. J. Physiol. Cell Physiol.* **289**, C1295–C1302 (2005).
73. Elliott, J. I. *et al.* Membrane phosphatidylserine distribution as a non-apoptotic signalling mechanism in lymphocytes. *Nature Cell Biol.* **7**, 808–816 (2005).
74. Chessell, I. *et al.* Disruption of the P2X<sub>7</sub> purinoceptor gene abolishes chronic inflammatory and neuropathic pain. *Pain* **114**, 386–396 (2005).
75. Ma, W. *et al.* Pore properties and pharmacological features of the P2X receptor channel in airway ciliated cells. *J. Physiol. (Lond.)* **571**, 503–517 (2006).
76. Yamamoto, K. *et al.* Impaired flow-dependent control of vascular tone and remodeling in P2X<sub>4</sub>-deficient mice. *Nature Med.* **12**, 133–137 (2006).
77. Agboh, K. C., Webb, T. E., Evans, R. J. & Ennion, S. J. Functional characterization of a P2X receptor from *Schistosoma mansoni*. *J. Biol. Chem.* **279**, 41650–41657 (2004).

**Acknowledgements** We thank the staff of the LMB Visual Aids Unit and J. A. Fisher for help with the illustrations. Work in our laboratories is supported by the MRC, EMBO, HFSP, NIH and Wellcome Trust. We regret that space limitations prevented us from citing many important original papers.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence should be addressed to B.S.K. ([bkhakh@mednet.ucla.edu](mailto:bkhakh@mednet.ucla.edu)) and R.A.N. ([r.a.north@manchester.ac.uk](mailto:r.a.north@manchester.ac.uk)).

# Dissecting self-renewal in stem cells with RNA interference

Natalia Ivanova<sup>1</sup>, Radu Dobrin<sup>1</sup>, Rong Lu<sup>1</sup>, Iulia Kotenko<sup>1</sup>, John LeVorse<sup>1</sup>, Christina DeCoste<sup>1</sup>, Xenia Schafer<sup>1</sup>, Yi Lun<sup>1</sup> & Ihor R. Lemischka<sup>1</sup>

**We present an integrated approach to identify genetic mechanisms that control self-renewal in mouse embryonic stem cells. We use short hairpin RNA (shRNA) loss-of-function techniques to downregulate a set of gene products whose expression patterns suggest self-renewal regulatory functions. We focus on transcriptional regulators and identify seven genes for which shRNA-mediated depletion negatively affects self-renewal, including four genes with previously unrecognized roles in self-renewal. Perturbations of these gene products are combined with dynamic, global analyses of gene expression. Our studies suggest specific biological roles for these molecules and reveal the complexity of cell fate regulation in embryonic stem cells.**

Mouse embryonic stem (ES) cells can differentiate into cells of all three germ layers, and self-renew extensively in culture<sup>1</sup>. The transcription factors *Oct4* (also known as *Pou5f1*) and *Nanog*, as well as the LIF–gp130–Stat3, BMP–TGF- $\beta$ –Smad, MAPK–ERK and possibly the WNT signalling pathways, all have important roles<sup>2–9</sup>. Cell fate choices in ES cells are clearly regulated by a complex orchestration of multiple pathways. Because this complexity is poorly understood, it has been difficult to produce specific lineages from ES cells and to re-programme somatic cell genomes after nuclear transplantation<sup>10,11</sup>. Here we dissect self-renewal using an integrated functional genomics strategy. We identify seven self-renewal regulators, suggest specific functional roles and provide a framework to analyse cell fate regulatory networks in ES cells.

## Identifying gene products required for self-renewal

We reasoned that like *Nanog* and *Oct4*, other genes required for self-renewal are downregulated upon induction of differentiation. Microarray time-course analyses of ES cells (six time points at 1-day intervals) undergoing retinoic-acid-induced<sup>12</sup> differentiation identified 901 rapidly downregulated genes (Supplementary Fig. 1). Sixty-five genes encoding transcription factors/DNA-binding proteins and unassigned expressed sequence tags (ESTs), as well as five additional retinoic-acid-insensitive genes identified in our previous study<sup>13</sup>, were chosen and downregulated using shRNA delivered by lentiviral vectors (Fig. 1a and Supplementary Fig. 2). The shRNA is expressed from an H1 promoter, while a constitutive ubiquitin C promoter drives hygromycin–green fluorescent protein (GFP) expression. More than 90% ES cell transduction is routinely achieved without drug selection. For a complete description see Supplementary Information.

To measure self-renewal in shRNA-transduced ES cells, we devised a fluorescence-based competition assay (Fig. 1b). GFP-positive (GFP<sup>+</sup>) shRNA-transduced ES cells were mixed in a 4-to-1 ratio with non-transduced GFP-negative (GFP<sup>−</sup>) cells, and cultured in the presence of LIF. The GFP<sup>+</sup>/GFP<sup>−</sup> ratios were measured at every passage. We anticipated that if a given shRNA induces ES cell differentiation, GFP<sup>+</sup>/GFP<sup>−</sup> ratios will decrease with time. Among the possible causes for such decrease are: (1) changes in cell cycle kinetics<sup>14</sup>; (2) changes in cell adhesion; and (3) compromised cell

survival due to the lack of specific growth factors that may be required for the maintenance of differentiating cells.

Progressive decreases in GFP<sup>+</sup>/GFP<sup>−</sup> ratios were observed upon downregulation of 10 out of 70 genes (Fig. 1c). These are: transcription factors *Nanog*, *Oct4*, *Sox2*, *Tbx3*, *Esrrb*<sup>2,4,15–17</sup>; a cofactor of the *Akt1* kinase *Tcl1* (ref. 18); an uncharacterized ES cell-specific gene *Dppa4* (ref. 11); and unassigned ESTs *Mm.343880*, *Mm.276044* and *Mm.219358*. Target messenger RNA levels were significantly decreased by all ten shRNAs. For detailed information see Supplementary Table 1.

Downregulation of some genes could retard the cell cycle without inducing differentiation. Accordingly, we evaluated cell morphology and alkaline phosphatase activity as indicators for undifferentiated cells. Downregulation of *Nanog*, *Oct4*, *Sox2*, *Tbx3*, *Esrrb*, *Tcl1*, *Dppa4* and *Mm.343880* resulted in morphological changes and loss of alkaline phosphatase activity (Supplementary Fig. 4). No colonies were obtained after downregulation of *Mm.276044* and *Mm.219358*, suggesting roles for these genes in the control of cell cycle or viability. In other cases, small alkaline-phosphatase-positive colonies were obtained and were not analysed further (data not shown). Morphological evaluation and alkaline phosphatase assays were performed for all 60 shRNA vectors that did not cause a change in GFP<sup>+</sup>/GFP<sup>−</sup> ratios. In no case was the morphology or alkaline phosphatase activity changed, indicating that the competition assay is a robust way to detect ES cell differentiation.

## Complementation rescue of compromised self-renewal activity

To rule out nonspecific effects of shRNAs we designed three additional shRNAs for each gene (Supplementary Table 1). At least one additional shRNA was effective for *Nanog*, *Oct4*, *Sox2*, *Esrrb* and *Tcl1*, showing that off-target effects are not responsible for the observed phenotypes. We also devised a genetic complementation (rescue) strategy where the original shRNA vectors were modified to include tetracycline-inducible versions of the shRNA-targeted gene products (Fig. 2a). To facilitate this rescue strategy the original shRNA targets were chosen to be in the 3' untranslated region (with the exception of *Oct4* and *Tcl1*). The shRNA-immune versions contain only the protein-coding sequences. Transactivator (rtTA)-expressing ES cells were transduced<sup>19</sup>. Self-renewal in transduced cells

<sup>1</sup>Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA.



is doxycycline-dependent, as shown in Fig. 2b for *Nanog*. *Nanog* rescue-vector-transduced GFP<sup>+</sup> cells (NanogR) were cloned in the presence of doxycycline and compared to clones transduced with empty vector (ControlR). Endogenous *Nanog* transcript was permanently downregulated, whereas vector-derived transcript was induced in the presence of doxycycline. The latter was undetectable 2 days after doxycycline removal. By morphological and alkaline phosphatase criteria, NanogR clones were undifferentiated in the presence of doxycycline. Upon doxycycline removal NanogR clones underwent morphological changes and lost alkaline phosphatase activity. Morphology and alkaline phosphatase activity of ControlR clones were unaffected by doxycycline (Fig. 2c). Robust doxycycline-dependent self-renewal was observed in *Sox2*, *Esrrb*, *Tbx3* and *Tcl1* rescue experiments. In the case of *Oct4*, we were not able to fully restore endogenous *Oct4* levels. Oct4R clones grow slowly but retain ES cell morphology, alkaline phosphatase activity and express high levels of *Nanog*. However, further analysis showed that several markers of trophoblast stem (TS) cells, such as *Hand1*, *Cdx2*, *Eomes* and *Mash2* (referenced in Supplementary Table 2), were upregulated in the presence of doxycycline (Supplementary Fig. 6). These cells probably represent an intermediate stage between ES and TS cells. After removal of doxycycline, Oct4R cells become negative for alkaline phosphatase, further upregulate TS markers and acquire characteristic trophoblast morphology. In the case of *Dppa4*, all rescue clones showed strong downregulation of endogenous mRNA but yielded a significant number of undifferentiated colonies (data not shown). Therefore, *Dppa4* may not be absolutely required for self-renewal. No rescue clones were obtained for *Mm.343880*, suggesting that the morphological changes observed in the initial shRNA experiments were potentially due to the off-target effects or technical

complications associated with vector packaging (Supplementary Fig. 5). This gene was excluded from most subsequent analyses. To confirm further the differentiated state of cells maintained without doxycycline we measured the expression levels of pluripotency markers *Nanog* and *Oct4*. After 8 days without doxycycline both markers were significantly reduced in NanogR, Sox2R, EsrrbR, Tbx3R and Tcl1R cells (Fig. 2e).

### Self-renewal regulators suppress differentiation

Loss of individual self-renewal regulators may initiate distinct differentiation programmes. We measured expression of early trophodermal, mesodermal, ectodermal and endodermal markers in the rescue clones during an 8-day period after doxycycline removal. In addition, the expression of markers for more mature cell types that originate from each germ layer was measured at day 8. In total, a set of 35 markers was analysed (Supplementary Table 2). All polymerase chain reaction with reverse transcription (RT-PCR) analyses were performed using two independent rescue clones. Observed marker patterns are shown in Supplementary Figs 6–9.

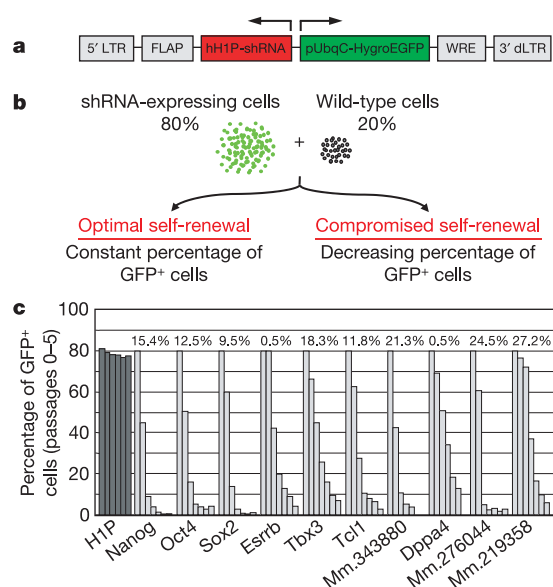
Trophodermal markers *Hand1*, *Cdx2*, *PL1*, *Mash2* and *Ehox* were induced in Oct4R, Sox2R and NanogR cells during the 8-day period after doxycycline removal. These markers were not induced when NanogR and Sox2R cells were maintained with doxycycline. The TS cell marker *Eomes* was upregulated only in Oct4R cells (Supplementary Fig. 6).

Goosecoid (*Gsc*) and brachyury (*T*), markers for primitive streak and early mesoderm, were transiently expressed in NanogR, Sox2R, EsrrbR and Tbx3R cells. In addition, *Mixl1*, expressed in cells of the organizer region that will become mesoderm, was induced in NanogR and Sox2R cells. After the expression of early mesodermal markers, expression of genes specific to mesoderm-derived cell lineages was detected. These are: *Cd34* and  $\beta$ -globin (extraembryonic mesoderm); *Bmp5*, *Shh* and *Tbx5* (dorsal mesoderm); *Gata5* and *Isl1* (cardiac mesoderm and definitive endoderm); *Meox1* and *Cart1* (lateral mesoderm). The induction of these markers was not observed when cells were maintained with doxycycline (Supplementary Fig. 7).

Expression of the primitive ectoderm marker *Fgf5* was transiently induced in NanogR, Sox2R, EsrrbR, Tbx3R and Tcl1R cells, and was followed by upregulation of the ectodermal marker *Cxcl12*. *Mash1* and *Pax3* but not *Sox1* were also induced, indicating that these cells differentiate into neural crest derivatives. Other neural crest markers such as *Ednra*, *Eya2*, *Ngn2*, *Phox2b* and *Slug* were also detected in these cells. Only a subset of these markers was induced in Tcl1R cells after doxycycline removal, suggesting that additional signals are required to sustain differentiation induced by downregulation of *Tcl1* (Supplementary Fig. 8).

We found that ControlR cells spontaneously upregulated endodermal markers when maintained without passaging for more than 6 days. Therefore, endodermal commitment was analysed 4 days after doxycycline removal. At this time endodermal markers (*Gata4*, *Gata6*, *Foxa2*, *Sox17*, *Nr2f1*, *Nr2f2* and *Bmp2*) were induced in NanogR, Sox2R and EsrrbR cells (Supplementary Fig. 9). The markers were induced at the same time or after the induction of the organizer gene *Gsc*. Therefore, these cells may be differentiating into definitive endoderm from *Gsc*<sup>+</sup>/*Foxa2*<sup>+</sup> precursors<sup>20</sup>. Collectively, our results demonstrate that in addition to both *Oct4* and *Nanog*, *Sox2*, *Esrrb*, *Tbx3* and *Tcl1* function to suppress ES cell differentiation *in vitro*. None of the differentiation programmes was induced after downregulation of *Dppa4*. We also confirmed that removal of doxycycline does not induce apoptosis or cell cycle arrest in any of the rescue cells (Supplementary Figs 17 and 18).

To explore possible interactions between the identified gene products and known signalling pathways implicated in the control of self-renewal, we asked whether the activation status of LIF, ERK1/2, BMP and WNT signalling pathways is altered by the shRNA treatments (Supplementary Fig. 12). Stat3, Smad1/5/8 and Smad2/3 phosphorylation as well as total levels of  $\beta$ -catenin were unaffected by any shRNAs.



**Figure 1 | Identification of self-renewal regulators in ES cells.** **a**, Lentiviral vectors expressing both a shRNA and a hygromycin-EGFP fusion protein were constructed for each of the selected genes. FLAP, nucleotide segment that improves transduction efficiency; dLTR, deleted long-terminal repeat; WRE, woodchuck hepatitis virus post-transcriptional regulatory element. **b**, A competition strategy was designed to identify gene products whose depletion compromises ES cell self-renewal. GFP<sup>+</sup> (shRNA-expressing) cells were mixed in a 4-to-1 ratio with untransduced (GFP<sup>-</sup>) cells, and cultured in the presence of LIF. With each passage the ratio of GFP<sup>+</sup>/total cells was measured by flow cytometry. **c**, Depletion of *Nanog* and *Oct4* as well as eight additional gene products results in impaired self-renewal. The proportions of GFP<sup>+</sup> cells transduced with the empty H1P vector do not change (representative experiment shown). In all cases Real-Time PCR (RT-PCR) confirms the reduction of mRNA levels (residual amounts relative to wild-type levels are shown as percentages above each time course).

ERK1/2 was hyper-phosphorylated after downregulation of *Nanog*, *Oct4*, *Sox2* and *Mm.343880*. The ERK1/2 pathway is activated in trophectoderm; therefore the increased phosphorylation of ERK1/2 correlates with the induction of the trophectodermal markers by *Nanog*, *Oct4* and *Sox2* shRNAs<sup>21,22</sup>.

### Constitutive expression of self-renewal regulators

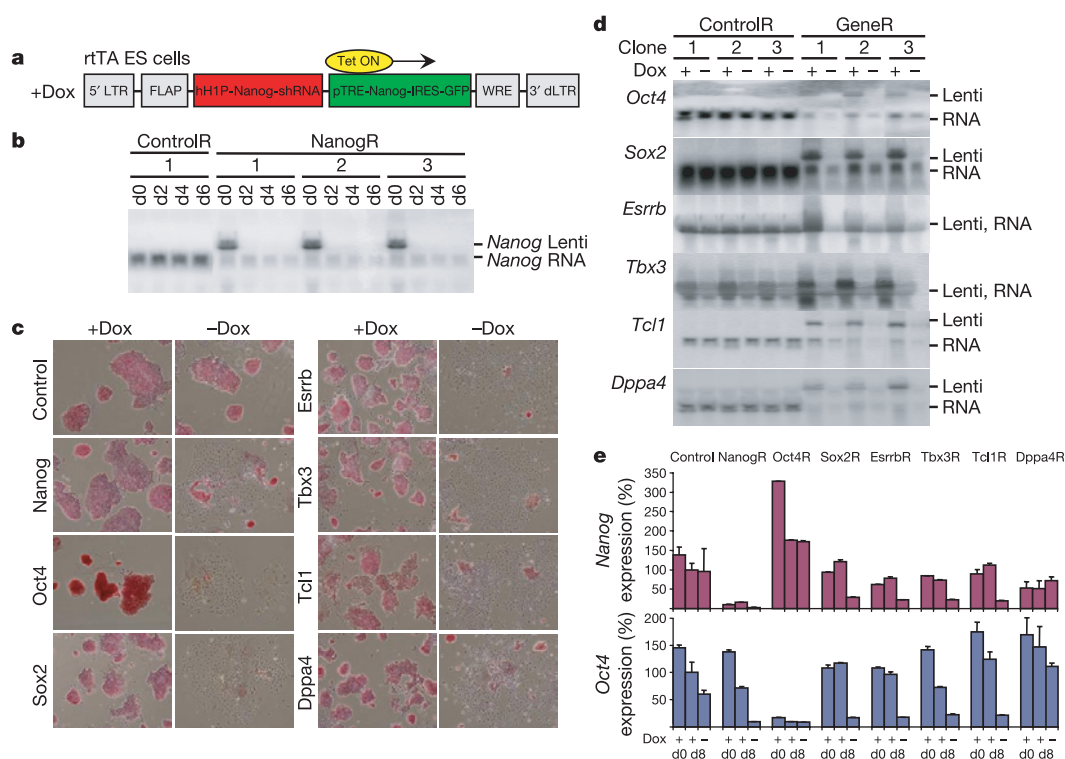
We next asked whether enforced expression of identified self-renewal regulators is sufficient to block the commitment to specific lineages. Rescue clones were used to produce embryoid bodies<sup>23</sup> in the presence of doxycycline. Temporal profiles of markers for neuroectoderm, mesoderm and endoderm were analysed during a 12-day period (Supplementary Fig. 10). Markers for all three lineages were induced in NanogR-, Tcl1R- and Dppa4R-derived embryoid bodies. Mesodermal commitment was affected in both Sox2R and Tbx3R cells. Marker patterns suggest that mesodermal development is blocked after the induction of *Gsc* in Sox2R cells (no *Mixl1* and no *T* expression) and slightly later in Tbx3R cells (normal *Mixl1* and no *T* expression). A complete block of both mesodermal and neuroectodermal differentiation was observed in EsrrbR cultures, where neuroectodermal differentiation was arrested after the induction of *Fgf5*, and mesodermal differentiation was blocked before the induction of *Gsc*. EsrrbR cells retained the capacity to form endoderm. In addition, expression of *Sox1* was increased in Sox2R cells and probably reflects the importance of *Sox2* in the neuroectodermal lineage<sup>24</sup>. Enforced expression of *Tcl1* may also promote the expansion of

neuroectodermal precursors and/or prevent their differentiation.

We also analysed the capacity of rescue cells to contribute to different tissues (brain, liver, muscle and coat hair) *in vivo*. ControlR cells contributed efficiently to all tissues, including the germ line, demonstrating that the rescue clone derivation procedures do not compromise pluripotency. NanogR, SoxR, Tbx3R and Tcl1R cells maintained without doxycycline for three passages did not contribute to any tissue, whereas EsrrbR cells retained limited capacity to contribute to neural crest derivatives such as coat melanocytes (see Supplementary Information for a detailed description).

### shRNA-induced gene expression dynamics

We next analysed transcriptome dynamics after downregulating the expression of each of the seven genes identified in the screen. shRNA-transduced GFP<sup>+</sup> cells were purified by fluorescence-activated cell sorting (FACS) daily during a 7-day interval and were used to interrogate Affymetrix microarrays. We identified a total of 3,109 non-redundant genes whose expression was perturbed after treatment with at least one shRNA (see Supplementary Information). Several distinct patterns of gene expression were observed. The first pattern (Fig. 3a) contains 771 genes that are up- or downregulated in response to most shRNA treatments. The second pattern (Fig. 3b) contains 474 genes perturbed in response to *Nanog*, *Oct4* and *Sox2* but not to *Esrrb*, *Tbx3*, *Tcl1* or *Dppa4* shRNA treatments. Recent chromatin studies have shown that *Nanog*, *Oct4* and *Sox2* often co-occupy the promoters of target genes<sup>25</sup>. The third pattern (Fig. 3c)



**Figure 2 | Genetic complementation to rescue shRNA-induced self-renewal defects.** **a**, The shRNA vectors were modified to produce tetracycline-inducible versions of endogenous targets. EGFP is coupled to the overexpressed gene via an internal ribosome entry sequence (IRES) (*Nanog* rescue vector is shown as an example). **b**, This approach was tested using *Nanog*. The rescue vectors were introduced into rtTA transactivator expressing ES cells in the presence of doxycycline (Dox); individual clones were isolated and expanded. Three clones (labelled 1–3) were chosen for further analysis. Control (ControlR) clones transduced with empty rescue vector (H1P-pTRE-IRES-EGFP) are also shown. Levels of endogenous *Nanog* (*Nanog* RNA) and vector derived *Nanog* (*Nanog* Lenti) were measured by northern blotting at 2-day intervals for 6 days (d0–d6) after withdrawal of doxycycline. In all three *Nanog*R clones endogenous *Nanog*

mRNA was permanently decreased. Vector-derived transcripts were expressed in the presence of doxycycline and rapidly reduced after doxycycline withdrawal. **c**, Self-renewal was rescued in the case of *Nanog* and for six other gene products. **d**, Levels of endogenous (RNA) and vector-derived (Lenti) transcripts were measured for three independent clones with and without doxycycline for the six additional genes. In all cases, the levels of endogenous transcript were decreased, and expression of the vector-derived transcripts was dependent on doxycycline. **e**, Expression of self-renewal markers *Nanog* and *Oct4* was reduced in the rescue cells maintained without doxycycline for 8 days. Expression level before the removal of doxycycline is shown as d0. Expression in d8 ControlR cells maintained with doxycycline is set as 100%. Each bar represents the average expression in three independent rescue clones corresponding to each gene; standard deviations are shown as error bars.



contains 272 gene products that are responsive to *Esrrb*, *Tbx3*, *Tcl1* and *Dppa4* shRNAs, but not to *Nanog*, *Oct4* or *Sox2* shRNAs. Patterns 2 and 3 suggest the existence of at least two distinct pathways that are necessary for self-renewal. Interestingly, whereas pattern 2 contains equal numbers of up- and downregulated gene products, pattern 3 contains gene products that are preferentially upregulated by shRNA treatments. This suggests that the corresponding pathway functions to repress differentiation-promoting genes. We have identified a number of smaller gene clusters that were perturbed by single shRNA treatments or by pairs of shRNA treatments (Supplementary Fig. 13).

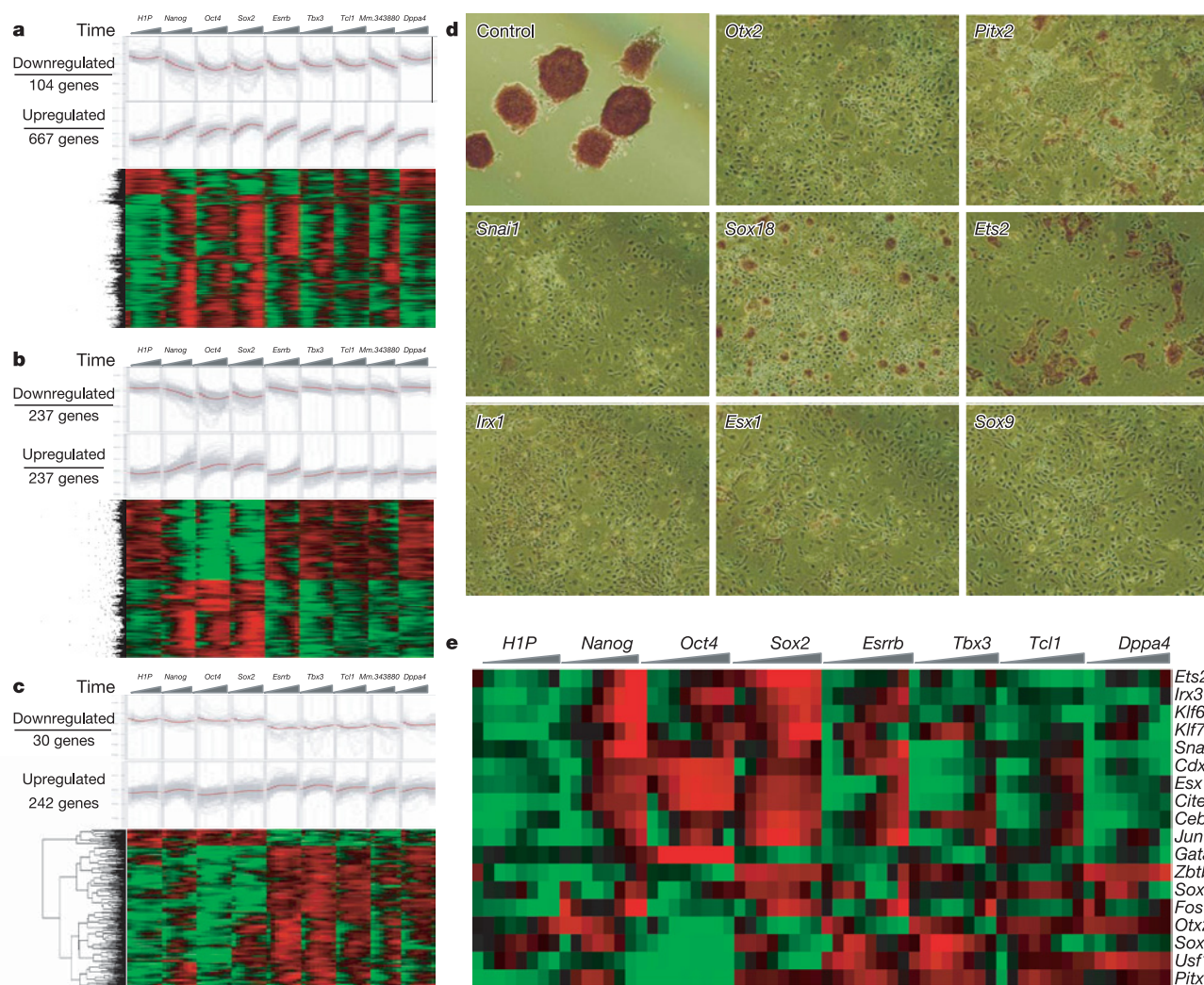
### Identification of gene products that promote differentiation

A number of transcriptional regulators are present in the gene expression patterns defined above. Some of these are upregulated after shRNA depletion of the self-renewal regulatory genes and may act as master positive regulators of differentiation. To identify such regulators, 160 gene products for which expression profiles suggest differentiation-inducing activity were overexpressed in ES cells. In 18 cases differentiation was suggested by changes in cell morphology,

loss of alkaline phosphatase activity and downregulation of *Nanog* (Fig. 3d and Supplementary Fig. 14). Temporal expression profiles of these 18 genes upon shRNA-mediated inactivation of self-renewal regulators (Fig. 3e) suggest that many differentiation inducers are under the control of multiple self-renewal regulators.

### *Nanog* compensates for loss of several self-renewal regulators

Most of the differentiation regulators identified in the overexpression screen are upregulated by *Nanog* shRNA treatment. *Nanog* is a strong positive regulator of self-renewal and counteracts removal of LIF and BMP<sup>4,7</sup>. We asked whether *Nanog* could block the differentiation induced by downregulation of the other genes identified in our studies. *Nanog* was overexpressed in ES cells as shown in Fig. 4a. This provides an approximately threefold increase in total *Nanog* levels and confers LIF independence (Supplementary Fig. 15). These cells were transduced with shRNA vectors, maintained in the presence of LIF, and evaluated using morphological criteria and alkaline phosphatase staining. *Nanog* overexpression restored undifferentiated morphology and alkaline phosphatase activity levels in *Esrrb*, *Tbx3*,



**Figure 3 | Global gene expression changes after downregulation of individual self-renewal regulators.** Microarray analysis time courses were performed to measure gene expression changes that accompany differentiation. **a–c**, Clustering revealed distinct patterns of gene expression. A number of transcriptional regulators present in the gene expression patterns that are upregulated upon shRNA-mediated depletion of self-renewal genes can induce differentiation when overexpressed in wild-type ES cells. **d**, Differentiation was apparent from the morphologies of the cells

as well as the loss of *Nanog* expression (data not shown). **e**, Temporal expression profiles of identified differentiation inducers upon shRNA-mediated inactivation of individual self-renewal regulators show that the genes (listed on the right side of the panel) are upregulated in at least one, and usually more than one, shRNA time course. Expression levels of the genes were centred and scaled to one standard deviation, and are indicated by colour, from green (low expression) to red (high expression).

*Tcl1* and *Dppa4* shRNA-transduced cells. Differentiation induced by depletion of *Oct4* and *Sox2* was unaffected (Fig. 4b and Supplementary Fig. 16). Therefore, the overexpression of *Nanog* may block differentiation into mesodermal, ectodermal and neural crest cell lineages. *Nanog* may block differentiation by simply upregulating the endogenous targeted gene to levels that are insensitive to shRNA depletion. However, in all cases the endogenous genes were similarly depleted both in *Nanog* overexpressing (NanogIP) and control (ControlIP) cells (Fig. 4c). Instead, the expression levels of other self-renewal regulators that are normally downregulated by shRNA treatments were fully or partially restored. Specifically, *Oct4* and *Sox2* were downregulated by *Esrrb* shRNA in ControlIP cells, but were maintained in NanogIP cells. *Nanog* overexpression restored the levels of *Oct4*, *Sox2* and *Esrrb* in *Tbx3* shRNA-treated cells and the levels of *Esrrb* and *Tbx3* in *Tcl1* shRNA-treated cells. Overexpression of *Nanog* was sufficient to prevent the induction of several differentiation inducers. *Cited1*, *Fos* and *Irx3* genes are shown in Fig. 4c as examples of such regulation.

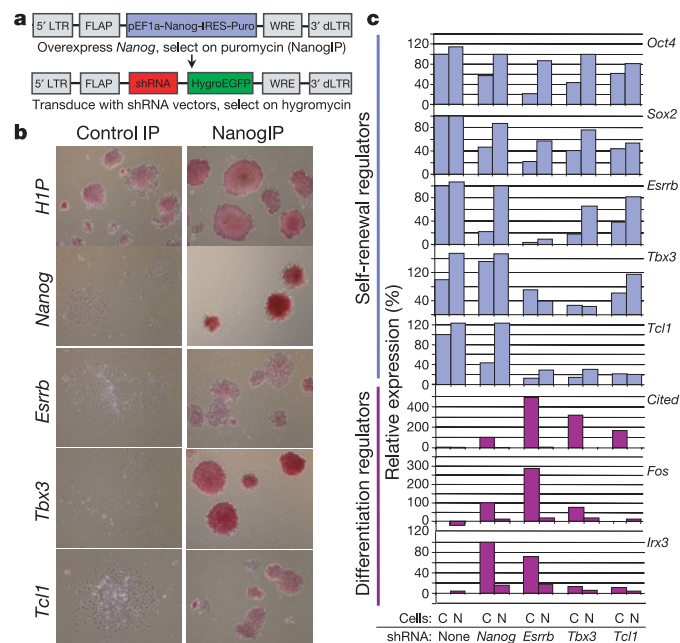
## Discussion

Using an integrated functional genomics approach we have demonstrated that *Esrrb*, *Tbx3* and *Tcl1*, as well as previously identified *Nanog*, *Oct4* and *Sox2*, are required for efficient self-renewal of ES cells *in vitro*. Downregulation of each gene induces differentiation of ES cells along specific lineages. The data obtained from our studies are summarized in Fig. 5.

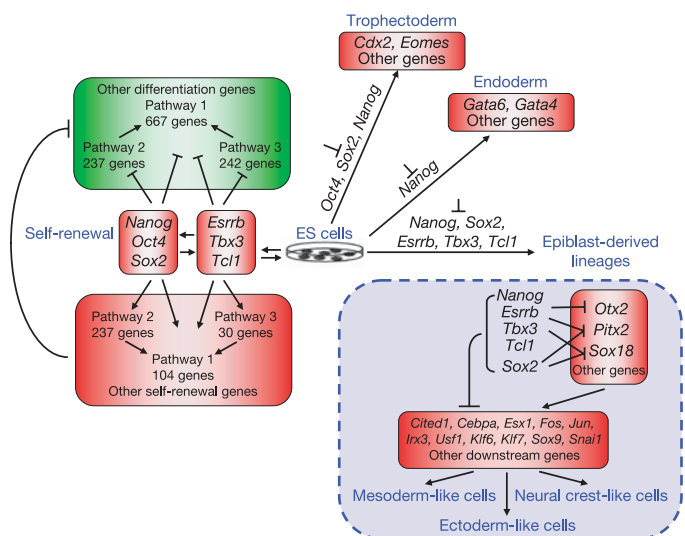
According to current models *Oct4* is required to prevent trophectodermal differentiation of ES cells<sup>3</sup>, whereas *Nanog* is required to block differentiation to the endodermal lineage<sup>5</sup>. Whereas *Oct4*

downregulation exclusively induces the set of trophectodermal genes, we found that, in addition to endoderm, *Nanog* downregulation induces the expression of markers for trophectoderm and epiblast-derived lineages, namely mesoderm, ectoderm and neural crest cells. Therefore, *Nanog* seems to be a global regulator that represses multiple differentiation programmes. *Sox2* functions to repress the development of trophectoderm and epiblast-derived lineages, whereas *Esrrb* and *Tbx3* are necessary to block the differentiation into mesoderm, ectoderm and neural crest cells but are not required to repress trophectoderm differentiation. The function of *Tcl1* seems to be even more restricted as it appears to repress only a subset of neural crest genes. In addition, enforced expression of *Sox2* and *Tbx3* is sufficient to repress the commitment to mesodermal lineage and *Esrrb* is capable of blocking both mesodermal and neuroectodermal commitment in embryoid body conditions.

We identify several patterns of transcriptional regulation. These patterns suggest the existence of two global pathways required for self-renewal and will be useful in efforts to further elucidate epistatic and other relationships. A number of genes that are upregulated after shRNA depletion of the self-renewal regulatory genes can act as positive regulators of differentiation programmes. Downregulation of *Nanog*, *Sox2*, *Esrrb*, *Tbx3* or *Tcl1* leads to the immediate induction of *Otx2*, *Pitx2*, *Sox18* and probably additional genes. All three genes are expressed in the epiblast; *Otx2* and *Pitx2* are important for mesodermal and neuroectodermal development *in vivo*<sup>26–29</sup>. *Otx2* seems to be a direct transcriptional target of both *Nanog* and *Oct4*, whereas *Pitx2* may be directly regulated by *Nanog* and *Sox2*, as suggested by recent chromatin-precipitation studies<sup>25,30</sup>. The activation of early-induced differentiation regulators is followed by the induction of the next wave of differentiation genes. Combinatorial action of these later-acting genes specifies the precise developmental outcome, such as mesoderm, ectoderm or neural-crest-like cells observed in our experimental settings. Notably, some of the late-induced genes seem to be direct transcriptional targets of self-renewal regulators. For instance, *Nanog* and *Oct4* are directly bound to the promoter regions of *Snai1*, *Klf6* and *Klf7* genes<sup>30</sup>.



**Figure 4 | *Nanog* rescue of phenotypes caused by shRNA directed against other self-renewal regulators.** **a**, ES cells were transduced with a *Nanog* overexpressing vector. *Nanog* overexpressing (NanogIP, N) and control (ControlIP, C) cells were challenged with the entire set of shRNA vectors. **b**, The cells were propagated in the presence of LIF, and analysed by morphology and alkaline phosphatase staining. **c**, Enforced *Nanog* activity does not simply upregulate the targeted endogenous genes to levels insensitive to shRNA inhibition but may also compensate for the loss of function of the targeted gene by maintaining the expression of other self-renewal regulators (blue) and by preventing the induction of key differentiation inducers (purple) that normally occurs upon shRNA-mediated downregulation of *Esrrb*, *Tbx3* or *Tcl1*. Gene expression levels were measured by microarrays.



**Figure 5 | Provisional model of cell fate regulatory interactions in ES cells *in vitro*.** *Esrrb*, *Tbx3* and *Tcl1*, as well as previously identified *Nanog*, *Oct4* and *Sox2*, are required for efficient self-renewal of ES cells *in vitro*. *Oct4* is required to prevent trophectodermal differentiation, *Nanog* and *Sox2* appear to be global regulators that repress multiple differentiation programmes, whereas *Esrrb*, *Tbx3* and *Tcl1* are necessary to block the differentiation into epiblast-derived lineages. Self-renewal regulators are integrated into an interconnected transcriptional network and control the expression of downstream target genes through distinct molecular pathways.



*Nanog* can substitute for *Esrrb*, *Tbx3* and *Tcl1* when overexpressed, probably through compensatory changes in the levels of other self-renewal regulators. This suggests that the self-renewal regulators are integrated into an interconnected transcriptional network and that the loss of function of one regulator can, in some cases, be compensated by adjusting the expression levels of other network components. Such a network would explain why knockouts of *Esrrb*, *Tbx3* and *Tcl1* show phenotypes at fairly advanced embryonic stages<sup>17,31–34</sup>. It will also be interesting to check whether *Sox2*, *Esrrb* and *Tbx3* can rescue differentiation into the epiblast-derived lineages induced by *Nanog* shRNA treatment. The function of *Dppa4* in ES cells remains unknown. None of the differentiation programmes analysed was induced in *Dppa4*R cells after removal of doxycycline, and the enforced expression of *Dppa4* did not affect differentiation in the embryoid body assay.

The screening strategy developed in this study can be extended to genome-scale shRNA libraries to identify additional gene products that have important roles in ES cell self-renewal<sup>35,36</sup>. It can also be applied to other systems, such as haematopoietic and neural stem cells, where developmental read-outs can be readily measured.

Regulatory similarities in mouse and human ES cells are unclear. *Oct4* and *Nanog* are essential; however, the LIF–gp130–Stat3 pathway appears not to be required in human ES cells<sup>37</sup>. It will be important to evaluate the panel of genes described herein as candidate regulators in the human ES system.

## METHODS

A detailed description of the methods used in these studies can be found in Supplementary Information.

**Cell lines.** Retinoic-acid-induced differentiation, competition assays, western blot analyses, microarray analyses of shRNA-induced differentiation and rescue of shRNA-induced phenotypes by *Nanog* overexpression were performed using mouse cell line CCE. Overexpression of shRNA-induced genes was performed using E14/T21 cells. E14/T21 cells were made by introducing EGFP under the control of the 6-kb *Nanog* promoter region into E14/T cells (a gift from A. Smith) using BAC transgenic technology (I.R.L. laboratory, unpublished data). CCE and E14/T21 ES cells were maintained on gelatin-coated dishes in DMEM media supplemented with 15% FBS (Hyclone), 100 mM MEM non-essential amino acids, 0.1 mM 2-mercaptoethanol, 1 mM L-glutamine (Invitrogen) and  $10^3$  U ml<sup>-1</sup> of LIF (Chemicon).

rtTA-expressing ES cells Aniv15 (a gift from G. Daley) were used for derivation of rescue clones, and were maintained on primary mouse embryonic fibroblasts in the above-described media supplemented with doxycycline (1 µg ml<sup>-1</sup>). For differentiation assays the cells were trypsinized and plated for 30 min on standard tissue culture dishes in order to remove primary mouse embryonic fibroblasts, and floating ES cells were collected and plated on gelatin-coated dishes.

Received 14 December 2005; accepted 16 May 2006.

Published online 11 June 2006.

- Smith, A. G. Embryo-derived stem cells: of mice and men. *Annu. Rev. Cell Dev. Biol.* **17**, 435–462 (2001).
- Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).
- Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24**, 372–376 (2000).
- Chambers, I. et al. Functional expression cloning of *Nanog*, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).
- Mitsui, K. et al. The homeoprotein *Nanog* is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
- Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* **12**, 2048–2060 (1998).
- Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of *Id* proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281–292 (2003).
- Burdon, T., Stracey, C., Chambers, I., Nichols, J. & Smith, A. Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev. Biol.* **210**, 30–43 (1999).
- Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A. H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature Med.* **10**, 55–63 (2004).

- Loebel, D. A., Watson, C. M., De Young, R. A. & Tam, P. P. Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Dev. Biol.* **264**, 1–14 (2003).
- Bortvin, A. et al. Incomplete reactivation of Oct4-related genes in mouse embryos cloned from somatic nuclei. *Development* **130**, 1673–1680 (2003).
- Tighe, A. P. & Gudas, L. J. Retinoic acid inhibits leukemia inhibitory factor signaling pathways in mouse embryonic stem cells. *J. Cell. Physiol.* **198**, 223–229 (2004).
- Ivanova, N. B. et al. A stem cell molecular signature. *Science* **298**, 601–604 (2002).
- Burdon, T., Smith, A. & Savatier, P. Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol.* **12**, 432–438 (2002).
- Avilion, A. A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126–140 (2003).
- Bamshad, M. et al. Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nature Genet.* **16**, 311–315 (1997).
- Mitsunaga, K. et al. Loss of PGC-specific expression of the orphan nuclear receptor ERR-β results in reduction of germ cell number in mouse embryos. *Mech. Dev.* **121**, 237–246 (2004).
- Pekarsky, Y. et al. Tc1 enhances Akt kinase activity and mediates its nuclear translocation. *Proc. Natl Acad. Sci. USA* **97**, 3028–3033 (2000).
- Ting, D. T., Kyba, M. & Daley, G. Q. Inducible transgene expression in mouse stem cells. *Methods Mol. Med.* **105**, 23–46 (2005).
- Yasunaga, M. et al. Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. *Nature Biotechnol.* **23**, 1542–1550 (2005).
- Corson, L. B., Yamanaka, Y., Lai, K. M. & Rossant, J. Spatial and temporal patterns of ERK signaling during mouse embryogenesis. *Development* **130**, 4527–4537 (2003).
- Saba-El-Leil, M. K. et al. An essential function of the mitogen-activated protein kinase Erk2 in mouse trophoblast development. *EMBO Rep.* **4**, 964–968 (2003).
- Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev.* **19**, 1129–1155 (2005).
- Ferri, A. L. et al. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* **131**, 3805–3819 (2004).
- Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
- Kurokawa, D. et al. Regulation of Otx2 expression and its functions in mouse epiblast and anterior neuroectoderm. *Development* **131**, 3307–3317 (2004).
- Essner, J. J., Branford, W. W., Zhang, J. & Yost, H. J. Mesendoderm and left-right brain, heart and gut development are differentially regulated by pitx2 isoforms. *Development* **127**, 1081–1093 (2000).
- Evans, A. L. & Gage, P. J. Expression of the homeobox gene Pitx2 in neural crest is required for optic stalk and ocular anterior segment development. *Hum. Mol. Genet.* **14**, 3347–3359 (2005).
- Zhang, C., Basta, T. & Klymkowsky, M. W. SOX7 and SOX18 are essential for cardiogenesis in *Xenopus*. *Dev. Dyn.* **234**, 878–891 (2005).
- Loh, Y. H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440 (2006).
- Luo, J. et al. Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-β. *Nature* **388**, 778–782 (1997).
- Davenport, T. G., Jerome-Majewska, L. A. & Papaioannou, V. E. Mammary gland, limb and yolk sac defects in mice lacking *Tbx3*, the gene mutated in human ulnar mammary syndrome. *Development* **130**, 2263–2273 (2003).
- Kang, S. M. et al. Impaired T- and B-cell development in *Tcl1*-deficient mice. *Blood* **105**, 1288–1294 (2005).
- Narducci, M. G. et al. *TCL1* participates in early embryonic development and is overexpressed in human seminomas. *Proc. Natl Acad. Sci. USA* **99**, 11712–11717 (2002).
- Berns, K. et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
- Paddison, P. J. et al. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
- Rao, M. Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Dev. Biol.* **275**, 269–286 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Smith for E14/T cells and the episomal expression system, G. Daley for Aniv15 cells, and D. Baltimore for the FUGW lentiviral expression system. We are grateful to K. A. Moore and members of the Lemischka laboratory for their assistance at various stages of these studies. We also thank A. Smith for constructive criticisms. This work was supported by funding from the NIDDK.

**Author Contributions** N.I. and I.R.L. designed the experiments. N.I., R.L., I.K., J.L., C.D., X.S. and Y.L. performed the experiments. N.I., R.D. and I.R.L. analysed the data. N.I. and I.R.L. wrote the paper.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.I. ([nivanova@molbio.princeton.edu](mailto:nivanova@molbio.princeton.edu)) or I.R.L. ([ilemischka@molbio.princeton.edu](mailto:ilemischka@molbio.princeton.edu)).

# The removal of cusps from galaxy centres by stellar feedback in the early Universe

Sergey Mashchenko<sup>1</sup>, H. M. P. Couchman<sup>1</sup> & James Wadsley<sup>1</sup>

The standard cosmological model, now strongly constrained by direct observations of the Universe at early epochs, is very successful in describing the evolution of structure on large and intermediate scales<sup>1</sup>. Unfortunately, serious contradictions remain on smaller, galactic scales<sup>1,2</sup>. Among the main small-scale problems is a significant and persistent discrepancy between observations of nearby galaxies, which imply that galactic dark matter haloes have a density profile with a flat core<sup>3–6</sup>, and the cosmological model, which predicts that the haloes should have divergent density (a cusp) at the centre<sup>7,8</sup>. Here we report numerical simulations that show that random bulk motions of gas in small primordial galaxies, of the magnitude expected in these systems, will result in a flattening of the central dark matter cusp on relatively short timescales ( $\sim 10^8$  years). Gas bulk motions in early galaxies are driven by supernova explosions that result from ongoing star formation. Our mechanism is general, and would have operated in all star-forming galaxies at redshifts  $z \geq 10$ . Once removed, the cusp cannot be reintroduced during the subsequent mergers involved in the build-up of larger galaxies<sup>9,10</sup>. As a consequence, in the present Universe both small and large galaxies would have flat dark matter core density profiles, in agreement with observations.

It is now widely accepted that structure in the Universe formed hierarchically, with small dark matter haloes forming first, later merging to make increasingly large virialized objects<sup>1</sup>. Analysis of cosmological simulations showed that dark matter haloes form with a central density cusp, with the innermost logarithmic slope being close to  $-1$  (refs 7, 8). This is in sharp contrast to observations which imply that galactic dark matter haloes have a flat central core<sup>3–6</sup>.

The proposed solutions to the problem of cosmological cusps can be broadly divided into three categories: (1) observational problems, (2) new physics (beyond standard cold dark matter (CDM) cosmology) and (3) conventional mechanisms (within the standard CDM cosmology). The observational solutions appear to be ruled out by the newest results, which suggest that neither limited angular resolution<sup>4</sup> nor non-circular gas motions<sup>5</sup> can be responsible for the flatness of the inferred galactic central density profiles. Among the new physics approaches, the warm dark matter scenario is problematic because the first galaxies capable of reionizing the Universe appear to form too late to explain the early reionization observed by the Wilkinson Microwave Anisotropy Probe<sup>11</sup>. Another modification of standard CDM, self-interacting dark matter, cannot simultaneously explain the very small cores observed in clusters of galaxies and the large cores observed in small galaxies<sup>12</sup>.

All conventional solutions within the standard CDM model rely on some source of gravitational heating of the dark matter to flatten the cusps. It has been suggested, for example, that a dark matter cusp can be erased by a central bar<sup>13</sup>, by passive gravitational evolution of self-gravitating gas clouds orbiting near the centre of the galaxy<sup>14</sup>, by recoiling black holes<sup>15</sup>, or when a powerful starburst drives all of the

gas out of the galaxy<sup>16</sup> ('maximum feedback'). These mechanisms are not believed to be sufficiently efficient or general, and are not widely accepted as resolving the discrepancy<sup>2</sup>.

We propose an efficient new mechanism for flattening the central dark matter cusp that is driven by the random bulk gas motions that are expected to be present in all star-forming primordial galaxies. Random bulk gas motions on scales of a few hundred parsecs are seen in all nearby large<sup>17,18</sup> and small<sup>19–21</sup> galaxies for which good quality radio observations are available, and are a predicted result of stellar feedback<sup>22–24</sup> (from the combined action of stellar winds and supernovae). In both observations and numerical simulations, the typical random velocities of the gas clouds are close to, or slightly above, the sound speed in interstellar gas, or  $\sim 10 \text{ km s}^{-1}$  (refs 17–24); clouds moving with highly supersonic speeds quickly lose kinetic energy via radiative shocks. Significantly, the spatial scale of the bulk gas motions is comparable to the scaling radius of high redshift ( $z \geq 10$ ) galaxies, and the gas clouds' velocities are close to their typical dark matter particle velocities. The result of the bulk gas motions is that the central gravitational potential of the galaxy fluctuates on a timescale comparable to the crossing time for dark matter particles, creating an efficient channel for transferring kinetic energy from gas to dark matter. The process is reminiscent of the violent relaxation that takes place in collapsing self-gravitating  $N$ -body systems<sup>25</sup>. That this effect has not been observed in numerical simulations to date is a result of the formidable computational challenge of simulating a realistic interstellar medium (ISM), subject to stellar feedback, inside a live dark-matter dwarf galaxy halo.

Our model is fundamentally different from those relying on passive gravitational evolution of gas clumps, in that it identifies a physically plausible source of energy to force hydrodynamic motion for long time periods. It is instructive to compare our model with that of ref. 14 in which the source of energy used to heat the dark matter is the orbital energy of massive, self-gravitating gas clouds, moving in the dense central part of the galaxy. This model correctly identifies moving massive clumps as the key mechanism able to transfer kinetic energy from gas to dark matter, but provides no mechanism to maintain these clouds or their kinetic energy on the timescales necessary to achieve cusp flattening. Indeed, such clouds would form stars and fragment owing to stellar feedback on a timescale of  $\sim 10^7$  years and, further, the supersonically moving clouds would slow down and fragment owing to the interaction with the surrounding inter-cloud medium on a comparable timescale. These timescales are much shorter than those required to flatten the cusp,  $> 10^8$  years. As a result, the energy of the orbital motion of gas clouds is not available for flattening the cusp. We propose instead that the natural source of energy for erasing the cusp is the energy of the stellar feedback. This energy is used to maintain the interstellar gas in the required clumpy state and to drive the gas clouds to sonic speeds over long intervals of time ( $> 10^8$  years); both effects are necessary if the cusp is to be removed.

<sup>1</sup>Department of Physics and Astronomy, McMaster University, Hamilton, Ontario, L8S 4M1, Canada.



**Table 1 | Model parameters**

| Parameter                            | Value                                |
|--------------------------------------|--------------------------------------|
| Halo virial mass, $m_{\text{vir}}$   | $10^9 M_\odot$                       |
| Halo virial radius, $r_{\text{vir}}$ | 3 kpc                                |
| Halo scaling radius, $r_s$           | 0.85 kpc                             |
| Halo scaling density, $\rho_s$       | $0.16 M_\odot \text{pc}^{-3}$        |
| Gas mass, $m$                        | $(0.25, 0.5, 1) \times 10^8 M_\odot$ |
| Radius of the gas clumps, $h$        | 40, 200 pc                           |
| Amplitude of the oscillations, $A$   | 400, 850 pc                          |
| Mean gas speed, $V$                  | $4\text{--}63 \text{ km s}^{-1}$     |
| Number of dark matter particles, $N$ | $32^3, 64^3, 128^3$                  |

The fiducial model corresponds to  $m = 10^8 M_\odot$ ,  $h = 40 \text{ pc}$  and  $A = 400 \text{ pc}$ .

To understand the essence of our mechanism, consider the central part of a dark matter halo with characteristic radius,  $R$ , and velocity dispersion,  $\sigma$ . If stellar feedback drives a significant fraction of the gas, of mass  $m$ , to one side of the system then the potential will fluctuate by an amount  $\sim Gm/R$  ( $G$  is Newton's constant). During one crossing time,  $\sim R/\sigma$ , dark matter particles can gain or lose kinetic energy of up to this amount depending on whether they are falling towards the gas concentration or moving away from it. If the random bulk velocity of the gas,  $V$ , is comparable to  $\sigma$ , subsequent motion of the gas concentration will reinforce a net increase in velocity dispersion,  $\sigma$ , and a secular transfer of kinetic energy to the dark matter. The effect is expected to be sensitive to the velocity dispersion of the gas bulk motions and the mass of the gas involved, but not to the density of the gas concentration. If the gas bulk motions are much faster than the velocity dispersion of dark matter particles ( $V \gg \sigma$ ), the effect becomes negligible, as the particles do not have time to react to changes in the potential and move instead in response to the time-averaged gas density distribution. In the opposite limit, with  $V \ll \sigma$ , the potential fluctuates slowly enough that the whole system can readjust adiabatically, again resulting in negligible energy transfer.

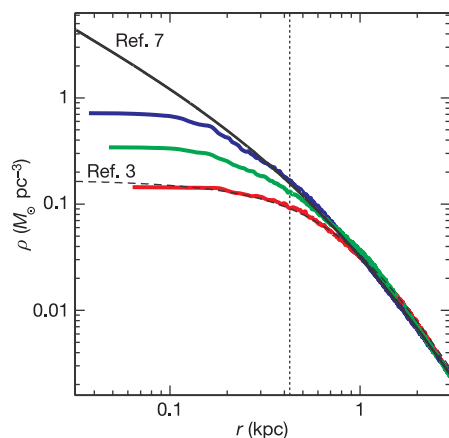
To estimate the importance of our mechanism in cosmological haloes, we ran a set of numerical simulations motivated by observational cues from nearby galaxies and that were designed to provide tractable physical models. Our model for the large-scale bulk gas motions in these systems places most of the gas in three large clumps moving as harmonic oscillators through the centre of the

galaxy along three orthogonal axes, with phases shifted by  $2\pi/3$  (see Table 1). This model provides the required imposed hydrodynamic forcing as well as permitting direct control over key parameters.

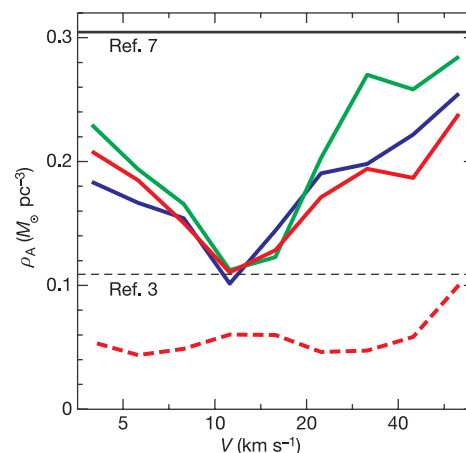
Figure 1 shows the evolution of the initial Navarro–Frenk–White<sup>7</sup> (NFW) density profile due to the model gas bulk motions (with  $V = 11 \text{ km s}^{-1}$ ), and shows that flattening of the central density cusp is effective on a timescale as short as 140 Myr, just one full period of the gas clump oscillation. At  $t = 140 \text{ Myr}$ , the dark matter density profile of the evolved halo is in remarkable agreement with observed (Burkert<sup>3</sup>) profiles. Thus the mechanism can operate on a timescale during which active stellar feedback would be expected in dwarf galaxies and on a sufficiently short timescale that the cusp flattening would be complete before the dwarf is involved in a significant merger in the structure formation hierarchy.

We tested the robustness and sensitivity of this result using several other parameter choices bracketing our fiducial model. Figure 2 shows the dependence of the cusp flattening on the mean speed,  $V$ , of bulk gas motions as well as on the gas mass and density. As expected, the most efficient transfer of kinetic energy from gas to dark matter occurs for gas velocities comparable to the velocities of dark matter particles: for our fiducial model, after  $\sim 200 \text{ Myr}$  of evolution, the effect is strongest (with the central dark matter density becoming comparable to that observed) for  $V = 10\text{--}20 \text{ km s}^{-1}$ . These velocities are in the same range as the observed velocity dispersions of interstellar gas bulk motions. After  $\geq 500 \text{ Myr}$  of evolution, a very significant cusp flattening is achieved for a much wider range of  $V$  (see Fig. 2; further simulations are described in Methods.)

We conclude that the observed properties of interstellar random bulk gas motions—a spatial scale of a few hundred parsecs and mildly supersonic—are just right to lead to fast erasure of the central dark matter cusp in small primordial galaxies. For very small galaxies ( $< 10^7 M_\odot$ , where  $M_\odot$  is the solar mass), the impact of stellar feedback is so strong that the whole ISM is driven out at supersonic speeds via galactic winds<sup>26</sup>. For much larger galaxies ( $> 10^{10} M_\odot$ ) stellar feedback plays a minor role, as the gravitational potential is deep enough to retain the ISM<sup>26</sup>, and the rotational speed of the disk is much larger than the speed of the random bulk gas motions. Our mechanism, in which stellar feedback drives most of the ISM in bulk motions that are confined within the dark matter halo, is hence expected to operate efficiently in a wide range of primordial galaxies



**Figure 1 | Evolution of the dark matter density profile in the fiducial model.** The dark matter density,  $\rho$ , is plotted as a function of the radius,  $r$ . The mean bulk gas speed is  $V = 11 \text{ km s}^{-1}$  and the number of dark matter particles is  $N = 128^3$ . The solid black line corresponds to the initial NFW<sup>7</sup> profile, the dashed black line shows the corresponding Burkert<sup>3</sup> profile. The blue, green and red lines show the density profiles of the simulated halo after 40, 80 and 140 Myr, respectively. The vertical dotted line marks the radius  $A = 400 \text{ pc}$  (the amplitude of the oscillations).



**Figure 2 | Degree of flattening of the central dark matter cusp for different models.** The averaged (within radius  $A = 400 \text{ pc}$ ) central dark matter density,  $\rho_A$ , is plotted as a function of the mean bulk gas speed,  $V$ . The black solid and dashed lines correspond to NFW<sup>7</sup> and Burkert<sup>3</sup> profiles, respectively. Red lines correspond to our fiducial model at  $t = 140 \text{ Myr}$  (solid) and  $t = 600 \text{ Myr}$  (dashed). The green line corresponds to the model with  $> 100$  times lower gas density ( $h = A/2 \approx 200 \text{ pc}$ ) at  $t = 220 \text{ Myr}$ . The blue line corresponds to the model with a two times smaller total gas mass at  $t = 240 \text{ Myr}$ .

whose masses lie somewhere between  $10^7 M_\odot$  and  $10^{10} M_\odot$ . During subsequent evolution, a fraction of the galactic gas will be consumed by star formation and lost via galactic winds, leaving a gas content in agreement with observations of present-day dwarf galaxies. As these galaxies merge together to make larger ones, the flat-cored shape of the dark matter density profile is preserved<sup>9,10</sup>. As a result, in the modern Universe most galaxies with masses  $\geq 10^8 M_\odot$  should have central dark matter densities smaller than the predictions of pure dark matter cosmological models—in agreement with observations.

A fully self-consistent demonstration of the effects described in this Letter requires the inclusion in simulations of detailed physical models of both star formation and feedback in high redshift galaxies. Ultimately, the goal is to follow at high resolution the hierarchical assembly of dwarf galaxy dark matter haloes in the full cosmological context together with the evolution of their gaseous and stellar contents.

It is noteworthy that if, indeed, most star-forming galaxies in the early Universe lost their dark matter cusps because of stellar feedback, another persistent cosmological problem could also be solved. The essence of this problem is that the standard cosmological model predicts that a large galaxy such as our own should have 10 to 100 times more small satellite galaxies than is observed<sup>27</sup>. Dwarf galaxies without a central cusp have a lower average core density than those with cusps, and are hence much easier to disrupt tidally during the hierarchical assembly of larger galaxies<sup>28</sup>. As a consequence, the removal of galactic cusps by stellar feedback in the early Universe would result in fewer satellite galaxies today.

## METHODS

**Model set-up.** We generated equilibrium dark matter haloes corresponding to a typical early Universe galaxy (see Table 1) with the NFW<sup>2</sup> density profile. The three gas clumps are represented by rigid spherical bodies with characteristic radius,  $h$ , mass,  $m$ , and softened gravitational acceleration  $g(r) = -Gm/(r^2 + h^2)$ . The spatial amplitude of the oscillations,  $A$ , was chosen to be  $r_s/2 \approx 400$  pc. All three bodies have the same mean speed,  $V$ , which is a free parameter. The gravitational  $N$ -body code GADGET<sup>29</sup> was used to evolve the models.

In our fiducial model, the total masses of gas and dark matter enclosed within the radius  $A$  are both equal to  $10^8 M_\odot$ . As the universal baryon to dark matter density ratio is  $\sim 1/5$  (ref. 11), this gas mass corresponds to about half of all gas in the galaxy. The radius,  $h$ , of the gas clumps is 40 pc. We explored a range of  $V$  from 4 to 63 km s<sup>-1</sup>, corresponding to  $(1/8-2)V_\odot$ , where  $V_\odot = 32$  km s<sup>-1</sup> is the circular velocity of the dark matter halo at  $r = A$ . The accuracy of the simulations was such that the total energy was conserved to better than 0.02% (for a dark-matter-only model).

**Model testing.** We tested our fiducial model with  $V = 11$  km s<sup>-1</sup> with three different resolutions ( $N = 32^3$ ,  $64^3$  and  $128^3$ ), and found the resulting dark matter density profiles of the evolved haloes to be identical within the measurement errors. We also re-simulated this model with much higher accuracy (with the resulting average time step being a factor of 7 shorter), and again found the results to be virtually identical to the original run.

We also checked if the initial adiabatic compression of the central part of the dark matter halo due to the presence of the gas would affect our results. Placing all of the gas ( $10^8 M_\odot$ ) in our fiducial model within the central 200 pc resulted in the central dark matter density slope becoming slightly steeper (logarithmic slope  $-1.4$  instead of  $-1$ ) after  $\sim 100$  Myr of evolution (no gas bulk motions were allowed). We then evolved this adiabatically compressed halo in the presence of bulk gas motions with parameters identical to our fiducial model with  $V = 11$  km s<sup>-1</sup>. After  $\sim 50$  Myr of evolution, the dark matter density profile became identical to that of the original run within the measurement errors.

A model with half the gas mass (corresponding to the averaged gas density within  $r = A$  being half that of the dark matter) still shows a very strong effect, with the time required to achieve complete cusp flattening being 240 Myr (see Fig. 2). Only when the averaged gas density drops to a quarter that of the dark matter does the effect become significantly weaker, with a cusp flattening time of  $\sim 800$  Myr (not shown).

In agreement with the discussion above, the strength of the effect does not depend significantly on the density of the gas clumps. In Fig. 2 we show the results for a model with  $h = A/2 \approx 200$  pc, with the gas clump density being

more than 100 times lower than in the fiducial case. In this model the Burkert-like central dark matter density is achieved after 220 Myr.

**Model comparisons.** The level of gas compression required by our model is not unreasonable, as cosmological simulations of dwarf galaxy formation show comparable or even larger central gas concentrations before the first starburst<sup>30</sup>. It is also consistent with the analytical prediction (which follows from the conservation of angular momentum) that the size of the galactic gaseous disk should be  $\sim \lambda r_{\text{vir}}$ , where  $\lambda \approx 0.05$  is the dimensionless angular momentum of cosmological haloes. Nevertheless, we tested the case of much lower gas concentration by running our model with the amplitude of oscillations,  $A$ , being twice as large ( $A = 0.85$  kpc), for the lower gas density case ( $h \approx 200$  pc). In this run, the time required to flatten the cusp becomes approximately twice as long (460 Myr), which is still smaller than the local Hubble age at redshift  $z = 10$ ,  $t_H = 490$  Myr. Interestingly, this time is again of the order of the oscillation period.

We also tested the effect of more than three gas clumps. For  $\geq 10$  clumps, we observed the central dark matter cusp becoming even more pronounced than initially (owing to adiabatic compression). This suggests that small-scale turbulence (with spatial scale  $\ll r_s$ ) cannot erase the central dark matter cusp.

The energy requirements for driving the bulk gas motions in our fiducial model with  $V = 11$  km s<sup>-1</sup> are quite modest. Ignoring energy input from stellar winds and assuming that each supernova releases  $10^{51}$  erg of thermal energy, we find a required supernova rate of  $8/\xi$  Myr<sup>-1</sup>, where  $\xi$  is the fraction of supernova energy transformed into kinetic energy of the bulk gas motions. Assuming a Salpeter stellar initial mass function and  $\xi = 0.1$ , this corresponds to a modest star formation rate of  $0.01 M_\odot \text{ yr}^{-1}$ . The corresponding gas depletion timescale is  $\sim 10$  Gyr, which is comparable to the values for the observed dwarf irregular galaxies.

Received 16 January; accepted 25 May 2006.

1. Coles, P. The state of the Universe. *Nature* **433**, 248–256 (2005).
2. Tasitsiomi, A. The state of the cold dark matter models on galactic and subgalactic scales. *Int. J. Mod. Phys. D* **12**, 1157–1196 (2003).
3. Burkert, A. The structure of dark matter halos in dwarf galaxies. *Astrophys. J.* **447**, L25–L28 (1995).
4. de Blok, W. J. G. & Bosma, A. High-resolution rotation curves of low surface brightness galaxies. *Astron. Astrophys.* **385**, 816–846 (2002).
5. Gentile, G., Burkert, A., Salucci, P., Klein, U. & Walter, F. The dwarf galaxy DDO 47 as a dark matter laboratory: testing cusps hiding in triaxial halos. *Astrophys. J.* **634**, L145–L148 (2005).
6. de Blok, W. J. G. Halo mass profiles and low surface brightness galaxy rotation curves. *Astrophys. J.* **634**, 227–238 (2005).
7. Navarro, J. F., Frenk, C. S. & White, S. D. M. A universal density profile from hierarchical clustering. *Astrophys. J.* **490**, 493–508 (1997).
8. Navarro, J. F. *et al.* The inner structure of  $\Lambda$ CDM haloes—III. Universality and asymptotic slopes. *Mon. Not. R. Astron. Soc.* **349**, 1039–1051 (2004).
9. Dehnen, W. Phase-space mixing and the merging of cusps. *Mon. Not. R. Astron. Soc.* **360**, 892–900 (2005).
10. Kazantzidis, S., Zentner, A. R. & Kravtsov, A. V. The robustness of dark matter density profiles in dissipationless mergers. *Astrophys. J.* **641**, 647–664 (2006).
11. Spergel, D. N. *et al.* First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: determination of cosmological parameters. *Astrophys. J. Suppl.* **148**, 175–194 (2003).
12. Sánchez-Salcedo, F. J. The dark halo of NGC 5963 as a constraint on dark matter self-interaction at the low-velocity regime. *Astrophys. J.* **631**, 244–251 (2005).
13. Weinberg, M. D. & Katz, N. Bar-driven dark halo evolution: a resolution of the cusp-core controversy. *Astrophys. J.* **580**, 627–633 (2002).
14. El-Zant, A., Shlosman, I. & Hoffman, Y. Dark halos: the flattening of the density cusp by dynamical friction. *Astrophys. J.* **560**, 636–643 (2001).
15. Merritt, D., Milosavljević, M., Favata, M., Hughes, S. A. & Holz, D. E. Consequences of gravitational radiation recoil. *Astrophys. J.* **607**, L9–L12 (2004).
16. Gnedin, O. Y. & Zhao, H. Maximum feedback and dark matter profiles of dwarf galaxies. *Mon. Not. R. Astron. Soc.* **333**, 299–306 (2002).
17. Brinks, E. & Bajaja, E. A high resolution hydrogen-line survey of Messier 31. III—H I holes in the interstellar medium. *Astron. Astrophys.* **169**, 14–42 (1986).
18. Deul, E. R. & van der Hulst, J. M. A survey of the neutral atomic hydrogen in M33. *Astron. Astrophys. Suppl.* **67**, 509–539 (1987).
19. Young, L. M. & Lo, K. Y. The neutral interstellar medium in nearby dwarf galaxies III. Sagittarius DIG, LGS 3, and Phoenix. *Astrophys. J.* **490**, 710–728 (1997).
20. Kim, S. *et al.* An H I aperture synthesis mosaic of the Large Magellanic Cloud. *Astrophys. J.* **503**, 674–688 (1998).
21. Begum, A. & Chengalur, J. N. Kinematics of the dwarf irregular galaxy GR8. *Astron. Astrophys.* **409**, 879–886 (2003).
22. Pelupessy, F. I., van der Werf, P. P. & Icke, V. Periodic bursts of star formation in irregular galaxies. *Astron. Astrophys.* **422**, 55–64 (2004).



23. Slyz, A. D., Devriendt, J. E. G., Bryan, G. & Silk, J. Towards simulating star formation in the interstellar medium. *Mon. Not. R. Astron. Soc.* **356**, 737–752 (2005).
24. de Avillez, M. A. & Breitschwerdt, D. Global dynamical evolution of the ISM in star forming galaxies. I. High resolution 3D simulations: Effect of the magnetic field. *Astron. Astrophys.* **436**, 585–600 (2005).
25. Lynden-Bell, D. Statistical mechanics of violent relaxation in stellar systems. *Mon. Not. R. Astron. Soc.* **136**, 101–121 (1967).
26. Mac Low, M.-M. & Ferrara, A. Starburst-driven mass loss from dwarf galaxies: efficiency and metal ejection. *Astrophys. J.* **513**, 142–155 (1999).
27. Moore, B. *et al.* Dark matter substructure within galactic halos. *Astrophys. J.* **524**, L19–L22 (1999).
28. Mashchenko, S. & Sills, A. Globular clusters with dark matter halos. II. Evolution in a tidal field. *Astrophys. J.* **619**, 258–269 (2005).
29. Springel, V., Yoshida, N. & White, S. D. M. GADGET: a code for collisionless and gasdynamical cosmological simulations. *New Astron.* **6**, 79–117 (2001).
30. Bromm, V. & Clarke, C. J. The formation of the first globular clusters in dwarf galaxies before the epoch of reionization. *Astrophys. J.* **566**, L1–L4 (2002).

**Acknowledgements** The simulations reported in this Letter were carried out on the McKenzie cluster at the Canadian Institute for Theoretical Astrophysics. H.M.P.C. is grateful for support from the Canadian Institute for Advanced Research, NSERC and SHARCNET. J.W. acknowledges support from NSERC.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.M. ([syam@physics.mcmaster.ca](mailto:syam@physics.mcmaster.ca)).

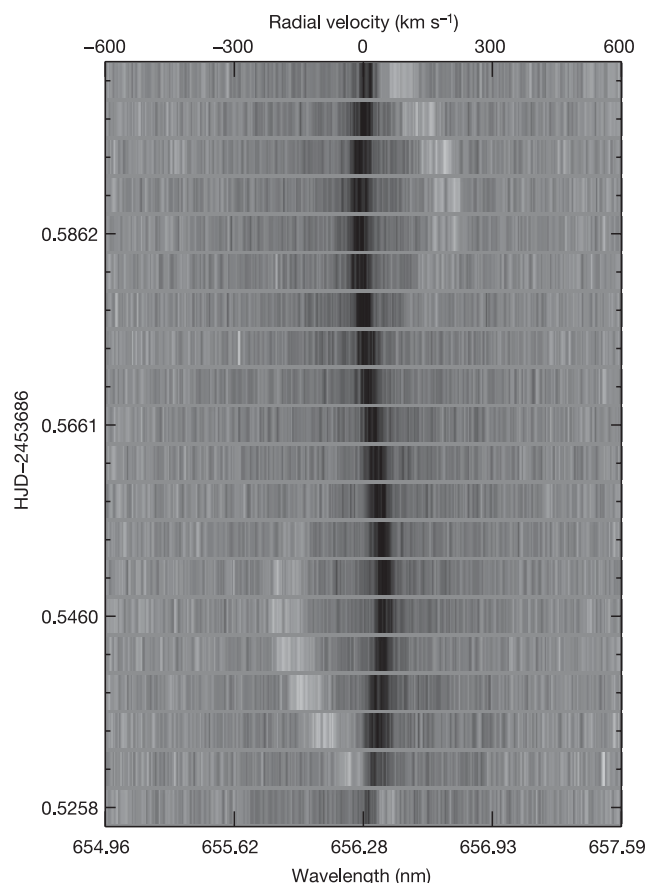
# Survival of a brown dwarf after engulfment by a red giant star

P. F. L. Maxted<sup>1</sup>, R. Napiwotzki<sup>2</sup>, P. D. Dobbie<sup>3</sup> & M. R. Burleigh<sup>3</sup>

Many sub-stellar companions (usually planets but also some brown dwarfs) orbit solar-type stars. These stars can engulf their sub-stellar companions when they become red giants. This interaction may explain several outstanding problems in astrophysics<sup>1–5</sup> but it is unclear under what conditions a low mass companion will evaporate, survive the interaction unchanged or gain mass<sup>1,4,5</sup>. Observational tests of models for this interaction have been hampered by a lack of positively identified remnants—that is, white dwarf stars with close, sub-stellar companions. The companion to the pre-white dwarf AA Doradus may be a brown dwarf, but the uncertain history of this star and the extreme luminosity difference between the components make it difficult to interpret the observations or to put strong constraints on the models<sup>6,7</sup>. The magnetic white dwarf SDSS J121209.31+013627.7 may have a close brown dwarf companion<sup>8</sup> but little is known about this binary at present. Here we report the discovery of a brown dwarf in a short period orbit around a white dwarf. The properties of both stars in this binary can be directly observed and show that the brown dwarf was engulfed by a red giant but that this had little effect on it.

WD 0137–349 (BPS CS 29504–0036) was first noted as an unremarkable faint, blue star in a survey for metal poor stars<sup>9</sup>. Improved spectroscopy showed it to be a white dwarf<sup>10</sup>. We obtained high resolution spectra of WD 0137–349 as part of the SPY programme to identify the progenitors of type Ia supernovae<sup>11</sup>. We noticed features in these spectra due to a low mass companion in a close orbit, so we obtained further observations (shown in Fig. 1) to measure the orbital period and mass ratio (Table 1). From the mass of the white dwarf (Table 2) and the mass ratio we derive a mass for the companion of  $(0.053 \pm 0.006)M_{\odot}$  (where  $M_{\odot}$  indicates solar mass). This is well below the limit of about  $0.075 M_{\odot}$  commonly used to distinguish stars from brown dwarfs<sup>12</sup>. Brown dwarfs, by definition, are not massive enough to support core hydrogen burning but do undergo a brief phase of deuterium burning soon after they form. They then start to cool, so the spectral type of a brown dwarf, which is a measure of its temperature, is also a measure of its age. The observed infrared flux distribution of WD 0137–349 (Fig. 2) is consistent with a model of an old brown dwarf companion with a mass of  $0.055 M_{\odot}$  but inconsistent with models for companions that are young brown dwarfs or stars. The orbital period of WD 0137–349 is approximately 116 min and the stars are separated by only  $0.65 R_{\odot}$  (where  $R_{\odot}$  indicates solar radius).

Brown dwarf companions to white dwarfs are rare—less than 0.5% of white dwarfs have a brown dwarf companion at any separation<sup>13</sup>. However, there are many white dwarfs that are known to have a low mass star as a companion in a short period orbit. In cataclysmic variable stars (CVs) mass transfer onto the white dwarf from the companion through the inner lagrangian point produces strong emission lines in the optical spectrum. No such lines are seen in



**Figure 1 | Triled spectrograms of WD 0137–349.** Spectra were obtained with the Ultraviolet–Visual Echelle Spectrograph (UVES) mounted on ESO’s Kuylen telescope. The radial velocity is indicated assuming a rest wavelength of 656.276 nm. The time of observation is indicated on the y axis. The exposure times are indicated by vertical extent of each spectrogram. The spectra have been normalized so the continuum value is 1. The grey-scale representation is a linear scale from 0.4 (black) to 1.4 (white). The sharp absorption feature (which appears black) is the H $\alpha$  line due to absorption by hydrogen in the atmosphere of the white dwarf star. The sinusoidal change in wavelength is due to the Doppler shift of the white dwarf as it orbits in a binary system. The emission feature seen moving in anti-phase to the absorption line arises in the atmosphere of the low mass companion to the white dwarf. The variation in the strength and width of this line shows that it is produced by irradiation of one hemisphere of the companion by the white dwarf. These observations are available from the ESO Science Archive Facility (programmes 276.D-5014 and 167.D-0407 at <http://archive.eso.org>). HJD, Heliocentric Julian Date.

<sup>1</sup>Astrophysics Group, Keele University, Keele, Staffordshire, ST5 5BG, UK. <sup>2</sup>Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK.

<sup>3</sup>Department of Physics and Astronomy, University of Leicester, University Road, Leicester, LE1 7RH, UK.

**Table 1 | Spectroscopic orbit of WD 0137-349**

|                                     |                           |
|-------------------------------------|---------------------------|
| $P$ (d)                             | $0.0803 \pm 0.0002$       |
| $T_0$ (Heliocentric Julian Date)    | $2453686.5276 \pm 0.0001$ |
| $K_1$ (km s $^{-1}$ )               | $27.9 \pm 0.3$            |
| $K_2$ (km s $^{-1}$ )               | $-187.5 \pm 1.1$          |
| $\gamma_1$ (km s $^{-1}$ )          | $17.8 \pm 0.3$            |
| $\gamma_2$ (km s $^{-1}$ )          | $3.4 \pm 1.0$             |
| Correction to $K_2$ (km s $^{-1}$ ) | $21 \pm 7$                |
| $m_1 \sin^3 i$ ( $M_\odot$ )        | $0.097 \pm 0.008$         |
| $m_2 \sin^3 i$ ( $M_\odot$ )        | $0.013 \pm 0.001$         |
| $a \sin i$ ( $R_\odot$ )            | $0.375 \pm 0.014$         |
| Mass ratio, $m_2/m_1 = K_1/K_2$     | $0.134 \pm 0.006$         |

The measured radial velocities of the H $\alpha$  absorption line at time  $T$  are given by  $\gamma_1 + K_1 \sin(2\pi(T - T_0)/P)$ , and similarly for the emission line. ( $P$  is orbital period,  $T_0$  is reference time,  $\gamma_1$  and  $\gamma_2$  are the apparent mean radial velocities,  $K_1$  and  $K_2$  are the semi-amplitudes of the spectroscopic orbits.) A correction to the value of  $K_2$  has been applied because the light in the emission line we measured is offset from the centre of the companion towards the centre-of-mass of the binary<sup>22</sup>. Accounting for this effect will increase the value of  $K_2$  by some fraction of the projected rotational velocity of the brown dwarf,  $v_{\text{rot}} \sin i$ , where  $i$  is the inclination of the orbital plane to the plane of the sky. The mass of the white dwarf is  $m_1$  and the mass of the brown dwarf is  $m_2$ . We estimate  $i \approx 35^\circ$  from the mass of the white dwarf given in Table 2 and the value of  $m_1 \sin^3 i$ . Strong tidal forces will ensure that the brown dwarf rotates synchronously, so for a typical brown dwarf radius of  $0.1 R_\odot$  we obtain  $v_{\text{rot}} \sin i = 35 \text{ km s}^{-1}$ . The minimum masses,  $m_1 \sin^3 i$  and  $m_2 \sin^3 i$ , and the projected separation ( $a \sin i$ ) are calculated using Kepler's laws from  $K_1$ ,  $K_2$  and  $P$ , including the correction to  $K_2$  described above. The difference  $\gamma_1 - \gamma_2 = 14.4 \pm 1.1 \text{ km s}^{-1}$  is due mainly to the gravitational redshift of the white dwarf, and agrees well with the value of  $13.3 \pm 1.1 \text{ km s}^{-1}$  implied by the parameters in Table 2.

our data for WD 0137-349, so we conclude that it is not a CV. There are several good candidates for sub-stellar companions in CVs, but the extra light due to accretion makes it difficult to confirm their masses<sup>14,15</sup>.

Short period white dwarf binaries with low mass companions that do not transfer mass are known as pre-CVs, because the loss of orbital angular momentum by gravitational wave radiation and other mechanisms will result in the shrinkage of the orbit, the initiation of mass transfer and the formation of a CV<sup>16</sup>. The timescale for WD 0137-349 to become a CV through the loss of gravitational wave radiation is about 1.4 Gyr, at which time the orbit period will be 60–80 min. This is close to the minimum orbital period seen in CVs. This raises the possibility that some fraction of CVs with very low mass companions may have formed as the result of the evolution of binaries like WD 0137-349, rather than by extensive mass loss from the companion. A simulation of the population of pre-CVs formed from binaries like WD 0137-349 shows that most of these binaries will have evolved from a solar-type star with a brown dwarf companion separated by a few astronomical units<sup>17</sup>. Although a few such systems have been found<sup>18</sup>, such binaries are known to be rare<sup>19</sup> so it is likely that the contribution of stars like WD 0137-349 to the total CV population is a few per cent or less.

The simulation of the population of binaries like WD 0137-349 assumes that white dwarfs with close, low mass companions are the result of 'common envelope evolution'. In this scenario, the more massive star in a binary system becomes a red giant once it has exhausted hydrogen in its core. The red giant will interact with its companion when its radius becomes comparable to the separation of the binary. The details of the interaction are uncertain but some low mass companions will be engulfed by the red giant—that is, the core of the red giant and the low mass companion share a common envelope. If the companion is not sufficiently massive to force the envelope to co-rotate with its orbit, the drag on the companion will cause it to quickly spiral in towards the core of the red giant. Some fraction of the orbital energy released,  $\alpha_{\text{CE}}$ , will be deposited as kinetic energy in the envelope, which is ejected from the binary system.

The radius of a red giant is determined principally by the mass of its core. This is effectively the mass of the resulting white dwarf, so we can calculate the value of  $\alpha_{\text{CE}}$  required to explain the formation of WD 0137-349 assuming a range of red giant masses,  $M_g$ . The lowest possible value of  $M_g$  is  $0.8 M_\odot$ , because stars less massive than this do not evolve to the red giant stage within the lifetime of the Galaxy. This

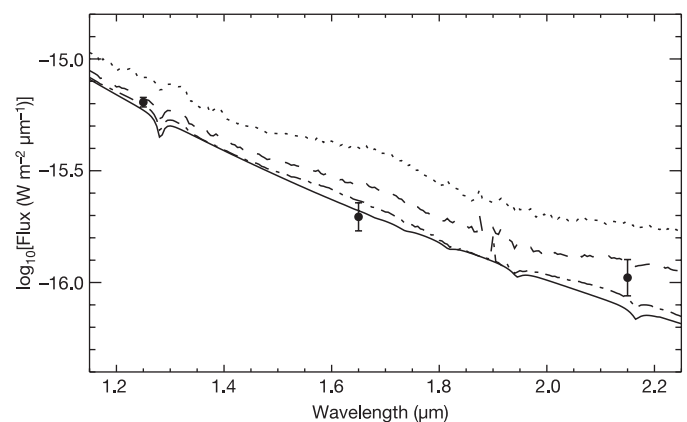
**Table 2 | Properties of the white dwarf WD 0137-349**

|                                             |                     |
|---------------------------------------------|---------------------|
| Apparent visual magnitude, $V$              | $15.33 \pm 0.02^9$  |
| Effective temperature, $T_{\text{eff}}$ (K) | $16,500 \pm 500$    |
| Surface gravity, $\log[g]$ (c.g.s. units)   | $7.49 \pm 0.08$     |
| Mass ( $M_\odot$ )                          | $0.39 \pm 0.035$    |
| Age (Myr)                                   | $250 \pm 80$        |
| Luminosity ( $L_\odot$ )                    | $0.023 \pm 0.004$   |
| Radius ( $R_\odot$ )                        | $0.0186 \pm 0.0012$ |
| Distance (pc)                               | $102 \pm 3$         |

The effective temperature ( $T_{\text{eff}}$ ) and surface gravity ( $g$ ) are derived from an analysis of the hydrogen absorption lines in the optical spectrum of WD 0137-349 using the technique described in ref. 23. The mass and age are inferred from  $T_{\text{eff}}$  and  $\log g$  using models of white dwarfs with a range of compositions<sup>24,25</sup>. The age is the time since formation of the white dwarf by ejection of the red giant envelope. The distance is inferred from the luminosity, apparent visual magnitude and a model atmosphere for hydrogen-rich white dwarfs<sup>26</sup>. The main factors that determine the white dwarf mass are well understood, and have been tested against observations, that is, pure hydrogen model atmospheres for moderately hot white dwarfs and the mass-radius relation for degenerate stars<sup>27</sup>.

provides a lower limit of  $\alpha_{\text{CE}} \approx 0.6$ . This value of  $\alpha_{\text{CE}}$  is similar to that derived for pre-CVs with more massive companions<sup>20</sup>. If the red giant was more massive than  $1.25 M_\odot$ , the value of  $\alpha_{\text{CE}}$  required exceeds 1. These values can be compared directly to the results of simulations of the common envelope phase, although no such simulations for systems resembling WD 0137-349 are available to us at present. Simple physical arguments suggest that low mass companions to red giants will be evaporated during the common envelope phase if they are less massive than some limit  $m_{\text{crit}}$ . The value of  $m_{\text{crit}}$  is uncertain, but is expected to be about  $0.02 M_\odot$  (refs 1, 4, 5). The properties of WD 0137-349 show that  $m_{\text{crit}}$  is at most  $(0.05\text{--}0.06)M_\odot$ . The mass of WD 0137-349 ( $0.4 M_\odot$ ) is lower than the typical mass of a single white dwarf ( $0.6 M_\odot$ ), as is expected for systems in which a common envelope phase has prematurely removed the envelope of a red giant<sup>21</sup>.

Some models predict that planets may accrete a substantial fraction of the mass in the red giant envelope before a common envelope phase, resulting in the formation of a binary with similar properties to WD 0137-349<sup>4</sup>. As most of the current mass of the brown dwarf was accreted from the red giant in this scenario, its spectral type would imply an age similar to that of the white dwarf—that is, about 250 Myr in the case of WD 0137-349. The spectral type in this case is expected to be  $\sim L1\text{--}2$  (effective temperature  $T_{\text{eff}} \approx 2,200 \text{ K}$ ). In contrast, if the brown dwarf has always been



**Figure 2 | Infrared flux distribution of WD 0137-349.** Measurements from the 2MASS archive (filled circles, with  $1\sigma$  error bars) are compared to a synthetic white dwarf spectrum from a pure-hydrogen model atmosphere normalized using the observed V-band magnitude (solid line). Also shown are the synthetic white dwarf spectrum combined with spectra of known brown dwarf stars scaled to the appropriate distance as follows (top-to-bottom): L0 (dotted line), L4 (dashed line) and T2 (dashed-dotted line)<sup>28</sup>. The orbital phase at which these data were obtained is unknown, so no account has been made for heating of the companion by the white dwarf.



close to its current mass, by the time the common-envelope phase occurs the brown dwarf will have been cooling for the lifetime of the solar-type star (gigayears). The common envelope phase proceeds on a dynamical timescale of a few years, which is negligible when compared to this thermal timescale, so very little mass or heat can be gained by the brown dwarf in this phase. The spectral type of the brown dwarf is then expected to be in the much cooler T-dwarf range ( $T_{\text{eff}} < 1,500 \text{ K}$ ). The companion may appear to be slightly hotter than this because it intercepts about 1% of the light from the white dwarf.

The asymmetric heating and rotation of the brown dwarf will produce a small but detectable orbital-period modulation of the brightness (at infrared wavelengths) of WD 0137–349; this is known as the ‘reflection effect’. An accurate measurement of the intrinsic spectral type and luminosity of the brown dwarf will therefore require infrared spectroscopy and photometry at a range of orbital phases to determine and account for the irradiation from the white dwarf. Despite this complication, the existing infrared photometry (Fig. 2) is more consistent with a spectral type for the brown dwarf that is slightly earlier (hotter) than T5, rather than with a spectral type of L1–2. Therefore, the existing data favour the model in which WD 0137–349 formed by a common envelope phase that had little effect on the brown dwarf, rather than by accretion onto a planet.

Received 24 February; accepted 13 June 2006.

- Nelemans, G. & Tauris, T. M. Formation of undermassive single white dwarfs and the influence of planets on late stellar evolution. *Astron. Astrophys.* **335**, L85–L88 (1998).
- Soker, N. & Harpaz, A. Rotation, planets, and the ‘second parameter’ of the horizontal branch. *Mon. Not. R. Astron. Soc.* **317**, 861–866 (2000).
- Harpaz, A. & Soker, N. Evaporation of brown dwarfs in AGB envelopes. *Mon. Not. R. Astron. Soc.* **270**, 734–742 (1994).
- Livio, M. & Soker, N. Star-planet systems as progenitors of cataclysmic binaries: tidal effects. *Astron. Astrophys.* **125**, L12–L15 (1983).
- Siess, L. & Livio, M. The accretion of brown dwarfs and planets by giant stars—II. Solar-mass stars on the red giant branch. *Mon. Not. R. Astron. Soc.* **308**, 1133–1149 (1999).
- Rauch, T. AA Dor—an eclipsing subdwarf - brown dwarf binary. *Rev. Mex. Astron. Astrophys. Ser. Conf.* **20**, 246 (2004).
- Hilditch, R. W., Kilkeny, D., Lynas-Gray, A. E. & Hill, G. New  $V_{\text{IC}}$  photometry of the sdOB binary AA Dor and an improved photometric model. *Mon. Not. R. Astron. Soc.* **344**, 644–650 (2003).
- Schmidt, G. D. et al. Discovery of a magnetic white dwarf/probable brown dwarf short-period binary. *Astrophys. J.* **630**, L173–L176 (2005).
- Preston, G. W., Shectman, S. A. & Beers, T. C. Photoelectric UVB photometry of stars selected in the HK objective-prism survey. *Astrophys. J. Suppl.* **76**, 1001–1031 (1991).
- Beers, T. C., Doinidis, S. P., Griffin, K. E., Preston, G. W. & Shectman, S. A. Spectroscopy of hot stars in the Galactic halo. *Astron. J.* **103**, 267–296 (1992).
- Napiwotzki, et al. Search for progenitors of supernovae type Ia with SPY. *Astron. Nachr.* **322**, 411–418 (2001).
- Liebert, J. W. Extrasolar planetary companions and brown dwarfs. In *Encyclopedia of Astronomy and Astrophysics* (ed. Murdin, P.) (<http://eaa.iop.org>) (2002); (Institute of Physics Publishing, Bristol, 2002).
- Farihi, J., Becklin, E. E. & Zuckerman, B. Low-luminosity companions to white dwarfs. *Astrophys. J. Suppl.* **161**, 394–428 (2005).
- Littlefair, S. P., Dhillon, V. S. & Martin, E. L. On the evidence for brown dwarf secondary stars in cataclysmic variables. *Mon. Not. R. Astron. Soc.* **340**, 264–268 (2003).
- Littlefair, S. P., Dhillon, V. S. & Martin, E. L. The K-band spectrum of the cataclysmic variable RXJ 0502.8+1624 (Tau 4). *Astron. Astrophys.* **437**, 637–639 (2005).
- Schreiber, M. R. & Gänsicke, B. T. The age, life expectancy, and space density of Post Common Envelope Binaries. *Astron. Astrophys.* **406**, 305–321 (2003).
- Politano, M. The formation of cataclysmic variable stars with brown dwarf secondaries. *Astrophys. J.* **604**, 817–826 (2004).
- Endl, M. et al. HD 137510: An oasis in the brown dwarf desert. *Astrophys. J.* **611**, 1121–1124 (2004).
- Grether, D. & Lineweaver, C. H. How dry is the brown dwarf desert? Quantifying the relative number of planets, brown dwarfs, and stellar companions around nearby Sun-like stars. *Astrophys. J.* **640**, 1051–1062 (2006).
- Nelemans, G. & Tout, C. A. Reconstructing the evolution of white dwarf binaries: further evidence for an alternative algorithm for the outcome of the common-envelope phase in close binaries. *Mon. Not. R. Astron. Soc.* **356**, 753–764 (2005).
- Marsh, T. R., Dhillon, V. S. & Duck, S. R. Low-mass white dwarfs need friends—Five new double-degenerate close binary stars. *Mon. Not. R. Astron. Soc.* **275**, 828–840 (1995).
- Maxted, P. F. L., Marsh, T. R., Moran, C., Dhillon, V. S. & Hilditch, R. W. The mass and radius of the M dwarf companion to GD448. *Mon. Not. R. Astron. Soc.* **300**, 1225–1232 (1998).
- Koester, D. et al. High-resolution UVES/VLT spectra of white dwarfs observed for the ESO SN Ia progenitor survey (SPY). *Astron. Astrophys.* **378**, 556–568 (2001).
- Benvenuto, O. G. & Althaus, L. G. Grids of white dwarf evolutionary models with masses from  $M = 0.1$  to  $1.2 M_{\odot}$ . *Mon. Not. R. Astron. Soc.* **303**, 30–38 (1999).
- Driebe, T., Schönberner, D., Blöcker, T. & Herwig, F. The evolution of helium white dwarfs. I. The companion of the millisecond pulsar PSR J1012RELSP+5307. *Astron. Astrophys.* **339**, 123–133 (1998).
- Bergeron, P., Wesemael, F. & Beauchamp, A. Photometric calibration of hydrogen- and helium-rich white dwarf models. *Publ. Astron. Soc. Pacif.* **107**, 1047–1054 (1995).
- Vauclair, G., Schmidt, H., Koester, D. & Allard, N. White dwarfs observed by the HIPPARCOS satellite. *Astron. Astrophys.* **325**, 1055–1062 (1997).
- Dobbie, P. D. et al. A near-infrared spectroscopic search for very-low-mass cool companions to notable DA white dwarfs. *Mon. Not. R. Astron. Soc.* **357**, 1049–1058 (2005).

**Acknowledgements** This Letter is based on observations collected at the European Southern Observatory, Chile, and makes use of data products from the Two Micron All Sky Survey (2MASS). P.D.D. is supported by PPARC. R.N. and M.R.B. acknowledge the support of PPARC Advanced Fellowships.

**Author Contributions** P.F.L.M. analysed and interpreted the data from which the orbital period and other properties of the system were measured. R.N. identified WD 0137–349 as a strong candidate for a brown dwarf companion to a white dwarf from observations obtained as part of the SPY programme, of which he is the principal investigator. M.R.B. and P.D.D. analysed the 2MASS data for this object. All authors discussed and interpreted the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.F.L.M. ([pflm@astro.keele.ac.uk](mailto:pflm@astro.keele.ac.uk)).

## LETTERS

# Interplay of electron–lattice interactions and superconductivity in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$

Jinho Lee<sup>1</sup>, K. Fujita<sup>1,2</sup>, K. McElroy<sup>1,3</sup>, J. A. Slezak<sup>1</sup>, M. Wang<sup>1</sup>, Y. Aiura<sup>4</sup>, H. Bando<sup>4</sup>, M. Ishikado<sup>2</sup>, T. Masui<sup>5</sup>, J.-X. Zhu<sup>6</sup>, A. V. Balatsky<sup>6</sup>, H. Eisaki<sup>4</sup>, S. Uchida<sup>2</sup> & J. C. Davis<sup>1</sup>

Formation of electron pairs is essential to superconductivity. For conventional superconductors, tunnelling spectroscopy has established that pairing is mediated by bosonic modes (phonons); a peak in the second derivative of tunnel current  $d^2I/dV^2$  corresponds to each phonon mode<sup>1–3</sup>. For high-transition-temperature (high- $T_c$ ) superconductivity, however, no boson mediating electron pairing has been identified. One explanation could be that electron pair formation<sup>4</sup> and related electron–boson interactions are heterogeneous at the atomic scale and therefore challenging to characterize. However, with the latest advances in  $d^2I/dV^2$  spectroscopy using scanning tunnelling microscopy, it has become possible to study bosonic modes directly at the atomic scale<sup>5</sup>. Here we report  $d^2I/dV^2$  imaging<sup>6–8</sup> studies of the high- $T_c$  superconductor  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ . We find intense disorder of electron–boson interaction energies at the nanometre scale, along with the expected modulations in  $d^2I/dV^2$  (refs 9, 10). Changing the density of holes has minimal effects on both the average mode energies and the modulations, indicating that the bosonic modes are unrelated to electronic or magnetic structure. Instead, the modes appear to be local lattice vibrations, as substitution of  $^{18}\text{O}$  for  $^{16}\text{O}$  throughout the material reduces the average mode energy by approximately 6 per cent—the expected effect of this isotope substitution on lattice vibration frequencies<sup>5</sup>. Significantly, the mode energies are always spatially anticorrelated with the superconducting pairing-gap energies, suggesting an interplay between these lattice vibration modes and the superconductivity.

Strong-coupling superconductivity theory<sup>1–3,11</sup>, allowed the impact of electron–phonon interactions on the superconducting density of states,  $\text{DOS}(E)$ , at energy  $E = \Delta + \Omega$  to be predicted ( $\Delta$  is the superconducting energy gap and  $\Omega$  the phonon energy). McMillan and Rowell used  $d^2I/dV^2$  measurements on planar normal-insulator-superconductor tunnel junctions<sup>1</sup> to reveal these signatures at boson energies  $\Omega = E - \Delta$ , identifying them with independently determined phonons of energy  $\Omega$ . The electron–phonon spectral function measured from these  $d^2I/dV^2$  spectra yielded the correct superconducting transition temperature—a milestone of twentieth-century physics.

By contrast, no consensus exists on the electron pairing mechanism of high- $T_c$  superconductivity. One reason is that the extensive studies of bosonic modes<sup>12–17</sup> and related electronic self-energy changes<sup>18–26</sup> have not resulted in the unambiguous identification of a boson mediating the pairing. Effects of electron–boson interactions (EBI) on electronic self-energies are most widely studied<sup>18–26</sup> via angle resolved photoemission. In the  $\Delta = 0$  or nodal direction of momentum space  $\mathbf{k} \parallel (\pi/a_0, \pi/a_0)$ , where  $a_0$  is the unit cell dimension, sudden changes or ‘kinks’ in quasiparticle dispersion  $E(\mathbf{k})$  occur

between 50 meV and 80 meV below the Fermi energy<sup>18–21</sup>. Several studies have addressed the issue of whether these ‘kinks’ might be due to magnetic interactions<sup>19,20</sup>. On the other hand, it has been proposed that, because of the doping independence of their energies, the ‘kinks’ are due to electron–phonon interactions<sup>21</sup>. In the antinodal directions  $\mathbf{k} \approx (\pi/a_0, 0)$  where  $\Delta$  is maximal, self energy changes<sup>22–25</sup> also occur between 50 meV and 90 meV below the Fermi energy. But here the effects are thought to occur at  $E = \Delta + \Omega$ . These modes have been discussed in terms of both magnetic interactions<sup>22–24</sup> and electron–phonon<sup>25</sup> interactions. Effects of  $^{16}\text{O}/^{18}\text{O}$  isotope substitution have been found primarily on states near  $\mathbf{k} \approx (\pi/a_0, 0)$ , providing evidence for electron interactions with lattice modes<sup>26</sup> (to be discussed in more detail below). Another EBI study technique is superconductor–insulator–superconductor (SIS) conductance measurement in break junctions<sup>27</sup>. Despite the challenge of interpreting SIS spectra in terms of the absolute value of  $\Omega$  (they are not a direct measure of the superconducting DOS<sup>28</sup>), clear EBI features are detected; they are analysed in terms of magnetic modes. A final probe of copper oxide EBI is optical spectroscopy<sup>29,30</sup>; it reveals self-energy changes that have been ascribed to magnetic interactions via a sharp mode<sup>29</sup> or a broad continuum<sup>30</sup>. Evidently, a definite conclusion for the identity of any pairing related EBI has proven elusive.

One reason for this situation might be that some key element of the EBI phenomenology has, so far, gone undetected. For example, if electron pairing<sup>4</sup> and any related EBI were disordered at the atomic scale, the techniques described above could not yield a complete description of the relevant nanoscale EBI phenomena because they average over space. For such reasons, atomic-resolution  $d^2I/dV^2$  spectroscopy of copper oxides has recently become the focus of considerable theoretical interest<sup>6–10</sup>. Three of the proposed applications are: (1) if the position ( $\mathbf{r}$ ) dependence of energy-gap disorder  $\Delta(\mathbf{r})$  were due to atomic-scale pair-potential disorder<sup>4</sup>, pairing-related EBI could be examined directly at that scale; (2) studies of specific local bosonic modes (unrelated to pairing) at defects or impurity/dopant atoms<sup>6–8</sup> could improve our understanding of EBI in copper oxides; and (3) Fourier transform  $d^2I/dV^2$  imaging (dubbed FT-IETS)<sup>10</sup> may help distinguish between different bosons involved in EBI. However, until now, severe technical constraints (see Supplementary Information) have prevented implementation of  $d^2I/dV^2$  imaging of the necessary spatial/energetic precision.

Here we report atomic-resolution  $d^2I/dV^2$ -imaging studies of a high- $T_c$  superconductor  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$  (Bi-2212). We use floating-zone-grown single crystals cleaved between the BiO planes in cryogenic ultrahigh vacuum and immediately inserted into the scanning tunnelling microscope head at 4.2 K. Figure 1a shows a

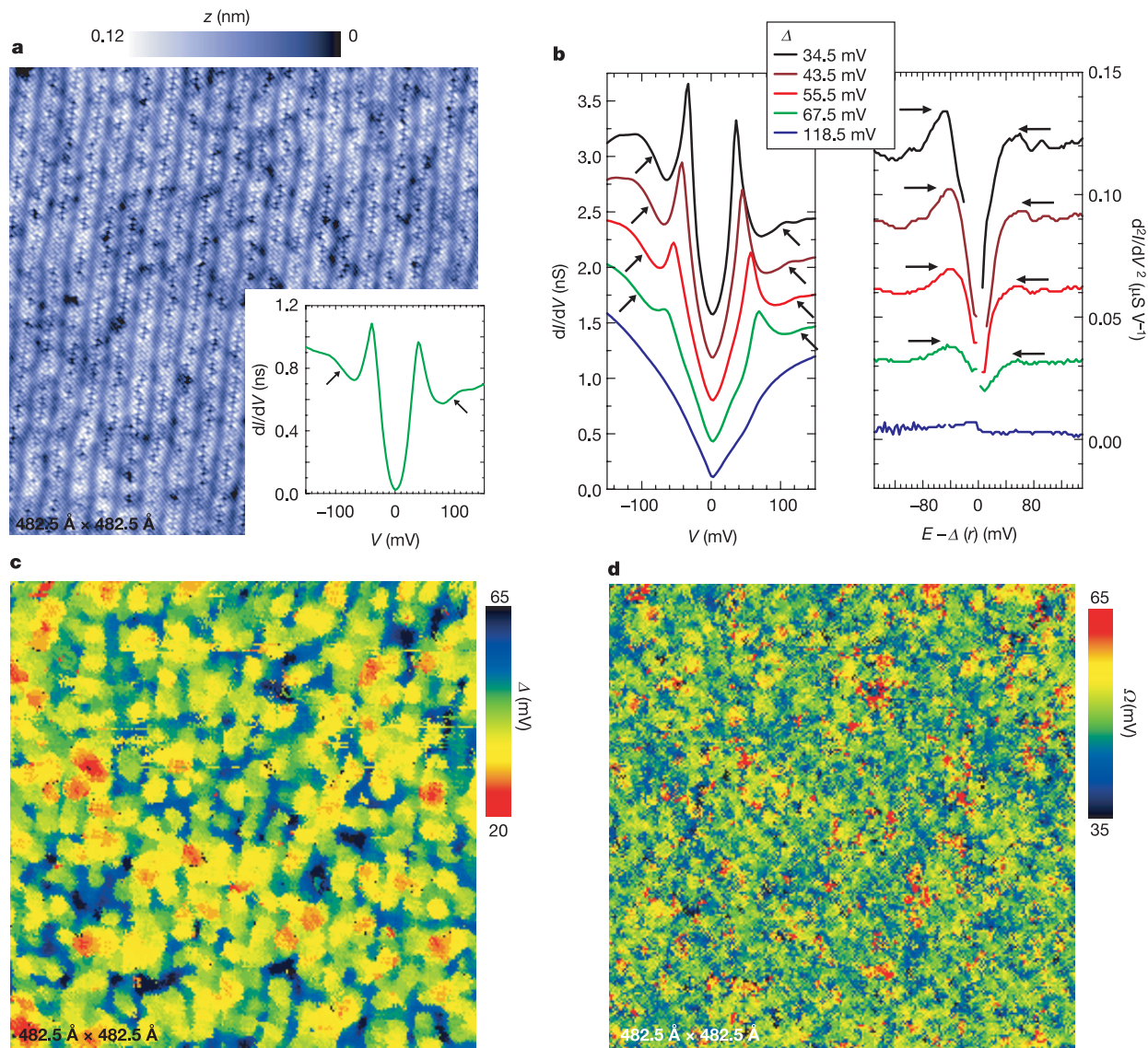
<sup>1</sup>LASSP, Department of Physics, Cornell University, Ithaca, New York 14853, USA. <sup>2</sup>Department of Physics, University of Tokyo, Tokyo, 113-8656, Japan. <sup>3</sup>Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. <sup>4</sup>AIST, 1-1-1 Central 2, Umezono, Tsukuba, Ibaraki, 305-8568, Japan. <sup>5</sup>Department of Physics, Osaka University, 1-1 Machikaneyama, Toyonaka, Osaka, 560-0043, Japan. <sup>6</sup>Theoretical Division, MS B262, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

topographic image of a typical BiO surface, while the inset shows the typical local density of states (LDOS) measured via a differential conductance  $dI/dV(E = eV)$  spectrum. Here we focus on the nano-scale spatial/energetic properties of the ubiquitous features at  $|E| > \Delta$  in the LDOS (for example, see arrows in Fig. 1a inset).

Figure 1b gives typical examples of  $dI/dV(E)$  spectra measured at different locations of Fig. 1a. The vast majority of these spectra exhibit peaks in  $d^2I/dV^2(E)$  occurring at the point of maximum slope in  $dI/dV$  for  $|E| > \Delta$  (arrows in Fig. 1b). Examples of the directly measured magnitude of the features in  $d^2I/dV^2$  are shown in Fig. 1b, where the horizontal axis has first been converted to  $\omega = E - \Delta$  (where  $\Delta$  is the distinct local gap magnitude for each spectrum). We provisionally consider these features as possible strong-coupling superconductivity<sup>1–3</sup> signatures of EBI (see Supplementary Information).

The  $dI/dV(\mathbf{r}, E)$  and  $d^2I/dV^2(\mathbf{r}, E)$  are simultaneously imaged

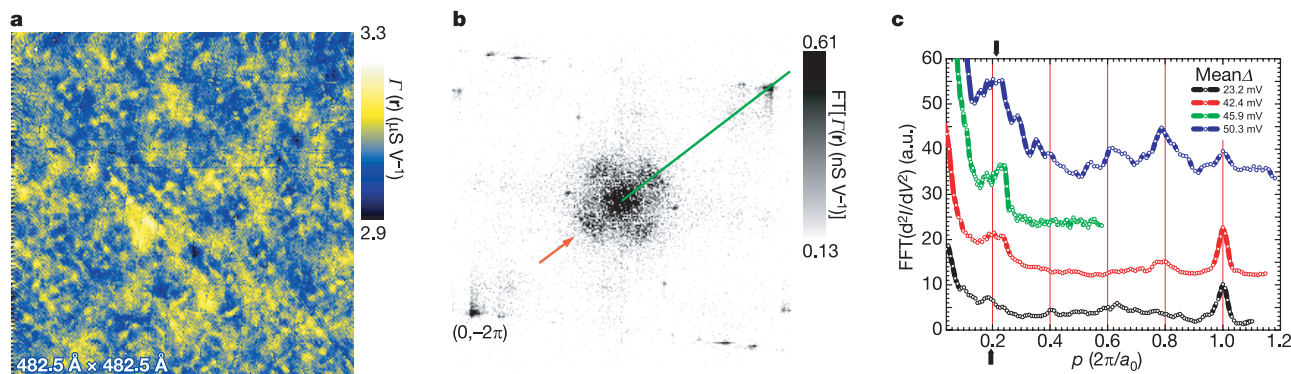
with atomic resolution and register. From the former, the gap-map  $\Delta(\mathbf{r})$  is derived (Fig. 1c). From the latter, the energies  $\Pi(\mathbf{r})$  at which peaks in  $d^2I/dV^2(\mathbf{r}, E)$  occur are measured. Within the context of strong-coupling superconductivity theory<sup>1–3,9</sup>, the local boson interaction energies would then be given by  $\Omega(\mathbf{r}) = \Pi(\mathbf{r}) - \Delta(\mathbf{r})$ . Some systematic uncertainties may exist in the precision with which this process yields the absolute mode energy (especially for complex band structures). Nevertheless, the mean mode energy is estimated from the  $\Omega(\mathbf{r})$  to be  $\bar{\Omega} \approx 52$  meV with a statistical spread of  $\pm 8$  meV. This is within the range of mode energies for antinodal EBI reported in photoemission studies<sup>22–25</sup>. However, it is from analysis of the  $\Omega(\mathbf{r})$  images (Fig. 1d, Fig. 3d–f) that a very different perspective on the EBI of Bi-2212 emerges. We see immediately that boson energies  $\Omega(\mathbf{r})$  are heterogeneous at the  $\sim 2$  nm scale with  $40 \text{ meV} < \Omega(\mathbf{r}) < 65 \text{ meV}$ , spanning the range of antinodal mode energies from photoemission<sup>22–25</sup>. The necessity of atomic-resolution  $d^2I/dV^2(\mathbf{r}, E)$ -imaging



**Figure 1 | Atomic-resolution  $d^2I/dV^2(\mathbf{r}, E)$  imaging of electron-boson interactions in Bi<sub>2</sub>Sr<sub>2</sub>CaCu<sub>2</sub>O<sub>8+δ</sub>.** **a**, Typical topographic image of one of the surfaces under study (Z, surface height). The inset shows the characteristic  $dI/dV$  spectrum, which is proportional to the DOS. The ubiquitous features occurring at energy  $E > \Delta$  (where  $\Delta$  is the superconducting energy gap), whose spatial and energetic structure are of central interest in this paper, are indicated by arrows. **b**, Examples of  $dI/dV$  spectra in different regions of surface shown in **a**; left, the peaks in  $d^2I/dV^2$  occur at the points of maximum slope of  $dI/dV$  for  $E > \Delta$  as indicated by arrows; right, examples of the directly

measured peaks in  $d^2I/dV^2$  from the identical locations as the same-coloured  $dI/dV$  spectra in **b**. **c**, The image of superconducting energy gap values, or gap-map  $\Delta(\mathbf{r})$ , on the surface in **a**. **d**, Image of the distribution of boson energies  $\Omega(\mathbf{r}) = \Pi(\mathbf{r}) - \Delta(\mathbf{r})$  (where  $\Pi(\mathbf{r})$  are bias energies at which  $d^2I/dV^2$  peaks occur) on the surface in **a**. This analysis scheme to find  $\Omega(\mathbf{r})$  has proven reliable and repeatable on numerous Bi-2212 samples at a wide range of dopings and with different oxygen isotopes. In all cases, it gives statistically indistinguishable results for both the filled and empty states:  $E < E_F$  and  $E > E_F$  ( $E_F$ , Fermi energy).





**Figure 2 | Ubiquitous quasi-periodic spatial modulations in  $d^2I/dV^2(r, \omega)$  signals, after gap referencing.** **a**, We define the local bosonic energy scale as  $\omega(r) = E - \Delta(r)$ . A typical map of  $d^2I/dV^2(r, \omega)$  modulations integrated over energy  $\Gamma(r) = \sum_{\omega=40 \text{ meV}}^{\omega=65 \text{ meV}} d^2I/dV^2(r, \omega)$  then reveals directly that the quasi-periodic spatial  $d^2I/dV^2(r, \omega)$  modulations are parallel to the Cu–O bond directions, have wavelengths of  $5a_0$  and correlation length of  $\sim 50 \text{ \AA}$ . These phenomena occur in all gap-referenced  $\Gamma(r)$ . **b**, The

characteristic wavevectors of these  $d^2I/dV^2(r, \omega)$  modulations, as indicated directly by the arrow in  $\Gamma(q)$ , the Fourier transform (FT) of  $\Gamma(r)$ , are  $\mathbf{p}_1 \approx 2\pi/a_0[(0.2, 0); (0, 0.2)] \pm 15\%$ . They are dispersionless within our resolution. **c**, The doping dependence of these  $\mathbf{p}_1$  are analysed by plotting the magnitude of  $\Gamma(q)$  along the line  $(0, 0)$  to  $(0, 2\pi)$ , as shown in **b**. We find that very similar modulation wavevectors (black arrows) occur at all dopings (FFT, fast Fourier transform; a.u., arbitrary units).

studies to fully explore EBI signatures in high- $T_c$  superconductors becomes manifest here.

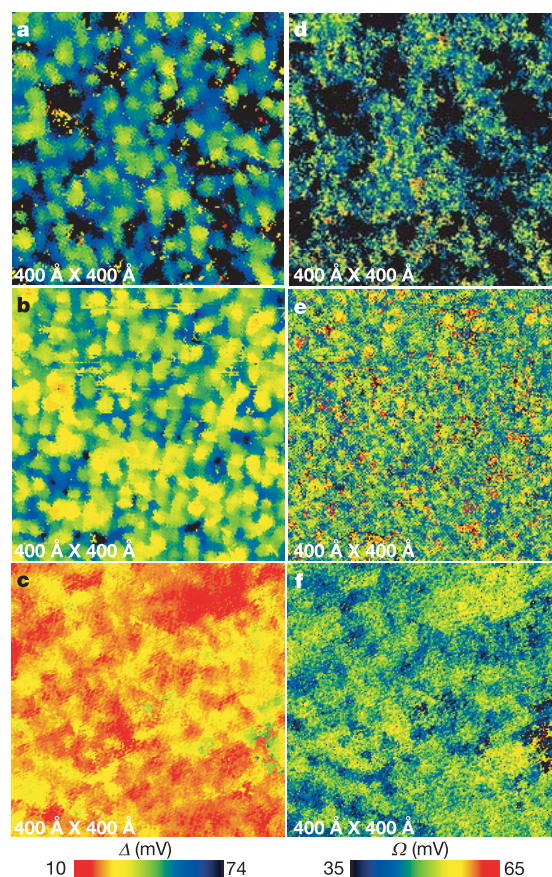
In theory<sup>9,10</sup>, spatial modulations of  $d^2I/dV^2(r, E)$  near  $E = \Omega + \Delta$  can contain key information about the EBI. But we detect no spatially periodic structure in unprocessed  $d^2I/dV^2(E)$  images in this energy range. This can be understood, however, because  $\Delta(r)$  is so strongly disordered (Fig. 1c) that any EBI<sup>10</sup> effects in  $d^2I/dV^2(r, \Omega + \Delta(r))$  would be spatially scrambled. To search for these effects, each  $d^2I/dV^2(E)$  must therefore be shifted to its bosonic energy scale  $\omega(r) = E - \Delta(r)$ , a process we refer to as ‘gap referencing’. This converts the unprocessed  $d^2I/dV^2(r, E); |E| > \Delta$  data into a new series of  $d^2I/dV^2(r, \omega)$  images. These vary little (except in intensity) within the energy range  $40 \text{ meV} < \omega < 65 \text{ meV}$  where peaks in  $d^2I/dV^2(\omega)$  are detectable (see Supplementary Fig. 1). Remarkably, they all exhibit the same spatial modulations, which did not exist in unprocessed  $d^2I/dV^2(r, E)$  data before gap referencing (see Supplementary Fig. 1).

To enhance their spatial contrast, we sum the  $d^2I/dV^2(r, \omega)$  images over the boson energy range where EBI are detected:  $\Gamma(r) = \sum_{\omega=40 \text{ meV}}^{\omega=65 \text{ meV}} d^2I/dV^2(r, \omega)$ . This energy average produces lower spatial noise compared to a single energy map. A typical resulting  $\Gamma(r)$  is shown in Fig. 2a; the modulations are parallel to the Cu–O bond directions and have wavelengths  $\sim 5a_0$  with correlation length of  $\sim 50 \text{ \AA}$ . We then use  $\Gamma(q)$ , the Fourier transform of  $\Gamma(r)$ , in Fig. 2b to determine that the modulation wavevectors are  $\mathbf{p}_1 \approx 2\pi/a_0[(0.2, 0); (0, 0.2)] \pm 15\%$ . Within the theoretical models<sup>9,10</sup>, such  $d^2I/dV^2(r, \omega)$  modulations are created when electronic states, renormalized by EBI, are scattered by disorder. When scattering is modelled as due to atomic-scale variations in the pair potential<sup>4</sup>, the predictions for  $d^2I/dV^2(r, \omega)$  modulations<sup>10</sup> are qualitatively consistent with data in Fig. 2, if the boson is a lattice vibration mode.

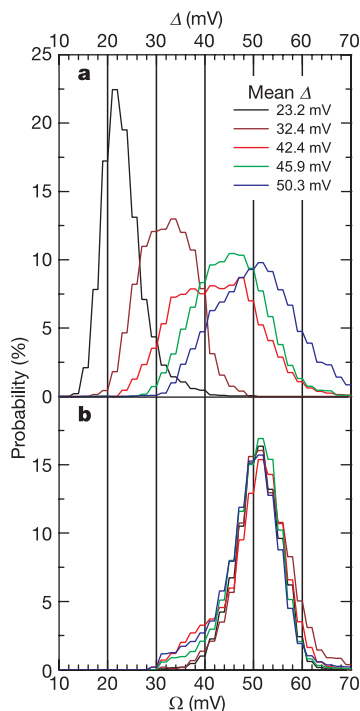
Next we study these  $d^2I/dV^2$  signatures of EBI at a sequence of different hole densities per  $\text{CuO}_2$ ,  $p$ :  $p \approx 0.12 \rightarrow p \approx 0.24$ . In Fig. 3a–c we show that the average superconducting energy gap  $\bar{\Delta}$  decreases from  $\sim 60 \text{ meV}$  to  $\sim 20 \text{ meV}$  with increasing doping, as expected. In strong contrast, Fig. 3d–f shows that, although changes occur in spatial correlations of  $\Omega(r)$ , no change is detectable in the average boson energy  $\bar{\Omega}$ . In fact, histograms of  $\Delta$  and  $\Omega$  measured on five samples at different dopings (Fig. 4) reveal that, while the distributions of  $\Delta$  evolve rapidly with doping, those of  $\Omega$  appear unchanged:  $\bar{\Omega} = 52 \pm 1 \text{ meV}$  for all dopings. We emphasize that the doping independence of  $\bar{\Omega}$  is not because the average energy  $\bar{\Omega}$  of the  $d^2I/dV^2$  peak is unchanged with doping;  $\bar{\Omega}$  changes from  $\bar{\Omega} \approx \pm 105 \text{ meV}$  to  $\bar{\Omega} \approx \pm 70 \text{ meV}$  over the doping range. It is the

difference,  $\bar{\Omega}$ , between  $\bar{\Omega}$  and  $\bar{\Delta}$  which remains constant. Furthermore, we find the  $d^2I/dV^2$ -modulation wavevectors  $\mathbf{p}_1$  are the same for all gap-referenced  $\Gamma(r)$  at all dopings (see Fig. 2c).

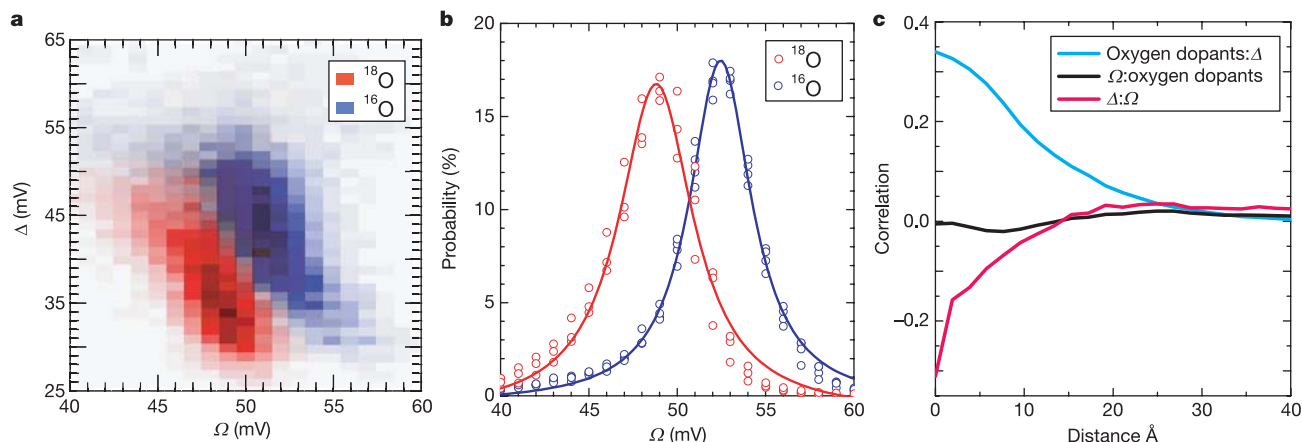
Thus, both  $d^2I/dV^2(r, \omega)$  modulations and  $\bar{\Omega}$  are independent of



**Figure 3 | Doping dependence of electron-boson interactions of  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ .** **a–c**, Gap-map  $\Delta(r)$  for three values of  $p$ , respectively  $\sim 0.12$ ,  $\sim 0.18$  and  $\sim 0.24$ . **d–f**, Simultaneously determined  $\Omega(r)$  images. Note that the colour bar is reversed here to show directly how higher  $\Omega$  is correlated to lower  $\Delta$ , and vice versa. Also, we can see that spatial correlations of  $\Delta(r)$  and  $\Omega(r)$  both change together with doping. Regions of **a** and **d** that are black are where neither the value of  $\Delta$  nor  $\Omega$  can be determined because there are no coherence peaks or  $d^2I/dV^2$ -peak features (see blue spectra in Fig. 1b).



**Figure 4 | Doping dependence of energy gap histograms and boson energy histograms in  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ .** **a**, Histograms of measured energy gaps  $\Delta$  from a sequence of samples with different dopings, black being strongly overdoped and blue strongly underdoped. We see clearly that  $\Delta$  falls rapidly with rising doping, and that the distribution of  $\Delta$  sharpens: there can be little doubt that the doping is indeed changing. **b**, Histograms of measured boson energies  $\Omega$ , from  $d^2I/dV^2$ -imaging measurements performed simultaneously with **a**. Within the uncertainty, neither the distribution nor the mean value of  $\Omega = 52 \pm 1$  meV are influenced by doping. Furthermore, since the same value of  $\Omega(r)$  is associated with different absolute values of  $\Delta(r)$  at different dopings, the most plausible explanation is that the doping-independent distributions of  $\Omega$  are inherent to the crystal. We note that in our  $>10^6$  atomically resolved  $dI/dV$  and  $d^2I/dV^2$  spectra in this study, the minimum in  $dI/dV$  always occurs near  $\Omega_{\text{dip}} = \pm 26$  meV at all dopings (see Supplementary Fig. 2).



**Figure 5 |  $^{18}\text{O}/^{16}\text{O}$  isotope effects on  $d^2I/dV^2(r, \omega)$  spectra and the distribution of boson energies.** The substitution of  $^{16}\text{O}$  by  $^{18}\text{O}$  was demonstrated via Raman spectroscopy of both in-plane and out-of-plane oxygen vibrational mode frequencies. **a**, Two dimensional  $\Delta\Omega$ -histograms of the frequency of occurrence of a given pair of  $\Delta, \Omega$  values in a single spectrum. Data for  $^{16}\text{O}$  are in blue and  $^{18}\text{O}$  in red. Although the  $\Omega$  vary in a fashion correlated with  $\Delta$ , the shift in  $\Omega$  with substitution of  $^{16}\text{O}$  by  $^{18}\text{O}$  is downwards by several meV. The vertical shift of 5.6 meV in  $\Delta$  between samples occurred inadvertently, because the hole density—as determined independently from  $T_c$  ( $^{16}\text{T}_c \approx 76$  K,  $^{18}\text{T}_c \approx 88$  K)—was slightly different

doped hole-density. Which boson<sup>13–17</sup> could exhibit such doping-independent EBI characteristics? The ‘resonant’ spin-1 magnetic excitation mode<sup>15</sup> appears inconsistent with the doping independence of mode energies because its energy is 43 meV in  $\text{Bi-2212}$  but, more importantly, is believed to be strongly doping dependent. The incommensurate, dispersive, spin density wave modes<sup>16,17</sup> also appear inconsistent because of their characteristic strong energy- or doping-dependences. By contrast, because energies of lattice-vibration modes change little with doping, they are logical candidates for the boson detected by  $d^2I/dV^2$  imaging.

Indeed, electron–lattice interactions are well known to influence copper oxide superconductivity. For example, the energy of a phonon mode at momentum transfer  $\mathbf{Q} \approx 2\pi/a_0[(0.25, 0); (0, 0.25)]$  diminishes rapidly towards 50 meV upon cooling into the superconducting state<sup>12–14</sup>, as might be expected for strong phonon interactions with superconducting quasiparticles. Furthermore, studies have revealed an unusual  $^{16}\text{O}/^{18}\text{O}$  isotope substitution effect<sup>26</sup> on the electronic structure; the data point to maximum influence of lattice modes on the high energy ( $E \approx -250$  meV) electronic structure near the antinodes, with a much weaker impact above  $T_c$  and at low energy. Although these results are not consistent with the simplest Eliashberg picture of electron–phonon interactions<sup>26</sup>, they do represent evidence for interactions between states near  $\mathbf{k} \approx (\pi/a_0, 0)$  and lattice vibrational modes. Furthermore, there have been detailed photoemission studies<sup>25</sup> of antinodal quasiparticles coupling to a bosonic mode—asccribed to a  $\text{B}_{1g}$  Cu–O bond-buckling phonon from theoretical analysis. Taken in combination, these studies point to interactions between antinodal quasiparticles and Cu–O-related lattice vibrations as influencing high temperature superconductivity—although the precise implications for electron pairing mechanism remain uncertain.

To test the hypothesis that bosons detectable by  $d^2I/dV^2$ -imaging techniques in  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$  are lattice vibration modes, we prepared crystals in which the normal  $^{16}\text{O}$  was completely substituted by  $^{18}\text{O}$  (as verified by frequency shifts detected in Raman spectroscopy). Figure 5a provides the comparisons between the distributions of  $\Omega(r)$  and  $\Delta(r)$  in different samples containing complete substitutions of the two oxygen isotopes. For each sample, we take  $\Delta(r)$  and  $\Omega(r)$  and construct a two-dimensional histogram of the frequency of occurrence of spectra with a given pair of values ( $\Delta, \Omega$ ). Each

in the two samples. As far as we know at present, no importance should be attributed to this shift. **b**, The histograms for all values of  $\Omega$  for samples with  $^{16}\text{O}$  (blue) and  $^{18}\text{O}$  (red). The average shift of energy with isotope substitution is  $-3.7 \pm 0.8$  meV. We find this same shift if  $\Omega$  distributions are measured for both filled ( $E > E_F$ ) and empty ( $E < E_F$ ) states. **c**, The normalized correlations between the dopant atom locations  $O(r)$  and both  $\Omega(r)$  and  $\Delta(r)$ . While zero-displacement dopant-gap-map  $O(r) : \Delta(r) \approx +0.35$  as expected, and  $\Omega(r) : \Delta(r) \approx -0.30$  consistent with Fig. 3, we find that  $\Omega(r)$  and  $O(r)$  are uncorrelated (solid black line).



histogram is peaked along the vertical axis at the most common gap energy  $\bar{\Delta}$  and along the horizontal axis at the most common boson energy  $\bar{\Omega}$ . Comparison between  $^{16}\text{O}$   $\Delta\Omega$ -histogram (blue) and the  $^{18}\text{O}$   $\Delta\Omega$ -histogram (red) reveals immediately that  $\bar{\Omega}$  for  $^{18}\text{O}$  shifts downwards by several meV compared to that of  $^{16}\text{O}$ . A quantitative analysis in Fig. 5b shows the distribution of boson energies  $\Omega$  in two different samples containing complete substitutions of the two oxygen isotopes:  $^{16}\text{O}$  in blue and  $^{18}\text{O}$  in red. We find that the shift of  $\bar{\Omega}$  upon substitution of  $^{16}\text{O}$  by  $^{18}\text{O}$  is  $-3.7 \pm 0.8$  meV. These results are found equally true for both filled  $E = -(\Delta + \Omega)$  and empty  $E = +(\Delta + \Omega)$  states, as expected for EBI in strong-coupling superconductivity theory<sup>2,3</sup>. Consequently, substitution of  $^{18}\text{O}$  for  $^{16}\text{O}$  reduces the mean boson energy scale  $\bar{\Omega}$  of EBI by 6% ( $\approx 1 - \sqrt{16/18}$ )—as expected for lattice vibrational modes involving the O atom.

These  $d^2I/dV^2$ -imaging studies alter several existing concepts of the EBI problem in  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ . We demonstrate that it is modes involving lattice vibrations that generate the  $d^2I/dV^2$  features in tunnelling. Further, the diminishing intensities of the  $d^2I/dV^2$  peaks with coherence peak height (Fig. 1b), along with the necessity for gap referencing (Fig. 2), signify the primary involvement of the antinodal states ( $E \approx \Delta$ ,  $\mathbf{k} \approx (\pi, 0)$ ) in the interactions. The mean mode energy is  $\bar{\Omega} \approx 52$  meV with a statistical spread of  $\pm 8$  meV. These modes exhibit intense atomic-scale disorder of interaction energies  $\Omega(\mathbf{r})$ , whose doping independence (Fig. 4b) points to a population native to the crystal. And perhaps most significantly, the  $\Omega(\mathbf{r})$  are spatially anticorrelated with the superconducting energy gap disorder  $\Delta(\mathbf{r})$  at all dopings (Fig. 3) and for both oxygen isotopes (Fig. 5a). Finally, in Fig. 5c we show the normalized correlations<sup>4</sup> between the dopant atom locations  $O(\mathbf{r})$  and both  $\Omega(\mathbf{r})$  and  $\Delta(\mathbf{r})$ . Whereas the zero-displacement correlations  $O(\mathbf{r}) : \Delta(\mathbf{r}) \approx +0.35$  are as expected, and the  $\Omega(\mathbf{r}) : \Delta(\mathbf{r}) \approx -0.30$  correlations are consistent with Fig. 3, we find that  $\Omega(\mathbf{r})$  and  $O(\mathbf{r})$  are uncorrelated. Therefore, correlations between  $\Omega(\mathbf{r})$  and  $\Delta(\mathbf{r})$  cannot be occurring trivially, through a similar effect of dopant disorder on both. A direct atomic-scale influence of  $\Omega(\mathbf{r})$  on  $\Delta(\mathbf{r})$  (or vice versa) is implied.

Taken together, these data present some intriguing new possibilities. The first is that superconducting energy gap disorder  $\Delta(\mathbf{r})$  is a consequence of heterogeneity in the pairing potential caused by disorder in the frequencies and coupling constants of pairing-related vibrational modes<sup>31</sup>. But the strong dependence of superconducting electronic structure on hole density while the  $\Omega(\mathbf{r})$  distributions remain unchanged (Fig. 4) appears to argue against this point of view. A second possibility is that the  $d^2I/dV^2$  features are unconnected to pairing-related EBI—perhaps occurring because of inelastic stimulation of vibrational modes within the tunnel barrier itself<sup>32</sup> or because of non-pairing-related electron lattice interactions. The primary difficulty here is that the ubiquitous anticorrelation between  $\Omega(\mathbf{r})$  and  $\Delta(\mathbf{r})$  cannot be explained trivially within such scenarios. A third possibility is that the  $d^2I/dV^2$  features represent electron–lattice interactions related to a competing electronic ordered state (see, for example, ref. 13), and that the anticorrelation between  $\Omega(\mathbf{r})$  and  $\Delta(\mathbf{r})$  occurs because of this competition. To help to distinguish between these possibilities, an atomic-scale version of the McMillan–Rowell procedure<sup>1</sup> may now become necessary.

Received 13 January; accepted 2 June 2006.

- McMillan, W. L. & Rowell, J. M. in *Superconductivity* Vol. 1 (ed. Parks, R. D.) 561 (Dekker, New York, 1969).
- Scalapino, D. J. in *Superconductivity* Vol. 1 (ed. Parks, R. D.) 449 (Dekker, New York, 1969).
- Carbotte, J. P. Properties of boson-exchange superconductors. *Rev. Mod. Phys.* **62**, 1027–1157 (1990).
- Nunner, T. S., Andersen, B. M., Melikyan, A. & Hirschfeld, P. J. Dopant-modulated pair interaction in cuprate superconductors. *Phys. Rev. Lett.* **95**, 177003 (2005).
- Stipe, B. C., Rezaei, M. A. & Ho, W. Single-molecule vibrational spectroscopy and microscopy. *Science* **280**, 1732–1735 (1998).
- Morr, D. K. & Nyberg, R. H. Localized bosonic modes in superconductors. *Phys. Rev. B* **68**, 060505 (2003).

- Han, J. H. Signature of a collective spin mode in the local tunnelling spectra of a *d*-wave superconductor. *Phys. Rev. B* **67**, 094506 (2003).
- Balatsky, A. V., Abanov, A. & Zhu, J.-X. Inelastic tunnelling spectroscopy in a *d*-wave superconductor. *Phys. Rev. B* **68**, 214506 (2003).
- Zhu, J.-X., Sun, J., Si, Q. & Balatsky, A. V. Effects of a collective spin resonance mode on the scanning tunnelling microscopy spectra of *d*-wave superconductors. *Phys. Rev. Lett.* **92**, 017002 (2004).
- Zhu, J.-X. *et al.* Effects of  $\tau_1$  scattering on Fourier-transformed inelastic tunnelling spectra in high- $T_c$  cuprates with bosonic modes. Preprint at (<http://www.arXiv.org/cond-mat/0507621>) (2005).
- Eliashberg, G. M. Interactions between electrons and lattice vibrations in a superconductor. *Sov. Phys. JETP* **11**, 696–702 (1960).
- McQueeney, R. J. *et al.* Anomalous dispersion of LO phonons in  $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$  at low temperatures. *Phys. Rev. Lett.* **82**, 628–631 (1999).
- Reznik, D. *et al.* Electron-phonon coupling reflecting dynamic charge inhomogeneity in copper oxide superconductors. *Nature* **440**, 1170–1173 (2006).
- Egami, T. Competing interactions in complex oxides: Polaron stability in manganites and cuprates. *AIP Conf. Proc.* **554**, 38–47 (2001).
- Sidis, Y. *et al.* Magnetic resonant excitations in high- $T_c$  superconductors. *Phys. Status Solidi B* **241**, 1204–1210 (2004).
- Tranquada, J. M. *et al.* Quantum magnetic excitations from stripes in copper oxide superconductors. *Nature* **429**, 534–538 (2004).
- Hayden, S. M., Mook, H. A., Dai, P., Perring, T. G. & Doğan, F. The structure of the high-energy spin excitations in a high-transition-temperature superconductor. *Nature* **429**, 531–534 (2004).
- Bogdanov, P. V. *et al.* Evidence for an energy scale for quasiparticle dispersion in  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ . *Phys. Rev. Lett.* **85**, 2581–2584 (2000).
- Kaminski, A. *et al.* Renormalization of spectral line shape and dispersion below  $T_c$  in  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ . *Phys. Rev. Lett.* **86**, 1070–1073 (2001).
- Johnson, P. D. *et al.* Doping and temperature dependence of the mass enhancement observed in the cuprate  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ . *Phys. Rev. Lett.* **87**, 177007 (2001).
- Lanzara, A. *et al.* Evidence for ubiquitous strong electron-phonon coupling in high-temperature superconductors. *Nature* **412**, 510–514 (2001).
- Norman, M. R. *et al.* Unusual dispersion and line shape of the superconducting state spectra of  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ . *Phys. Rev. Lett.* **79**, 3506–3509 (1997).
- Borisenko, S. V. *et al.* Anomalous enhancement of the coupling to the magnetic resonance mode in underdoped  $\text{Pb-Bi2212}$ . *Phys. Rev. Lett.* **90**, 207001 (2003).
- Gromko, A. D. *et al.* Mass-renormalized electronic excitations at  $(\pi, 0)$  in the superconducting state of  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ . *Phys. Rev. B* **68**, 174520 (2003).
- Cuk, T. *et al.* Coupling of the  $\text{B}_{1g}$  phonon to the antinodal electronic states of  $\text{Bi}_2\text{Sr}_2\text{Ca}_{0.92}\text{Y}_{0.08}\text{Cu}_2\text{O}_{8+\delta}$ . *Phys. Rev. Lett.* **93**, 117003 (2004).
- Gweon, G.-H. *et al.* An unusual isotope effect in a high-transition-temperature superconductor. *Nature* **430**, 187–190 (2004).
- Miyakawa, N. *et al.* Predominantly superconducting origin of large energy gaps in underdoped  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$  from tunnelling spectroscopy. *Phys. Rev. Lett.* **83**, 1018–1021 (1999).
- Romano, P. *et al.* Modelling study of the dip/hump feature in  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$  tunnelling spectroscopy. Preprint at (<http://www.arXiv.org/cond-mat/0602405>) (2006).
- Carbotte, J. P., Schachinger, E. & Basov, D. N. Coupling strength of charge carriers to spin fluctuations in high-temperature superconductors. *Nature* **401**, 354–356 (1999).
- Hwang, J., Timusk, T. & Gu, G. D. High-transition-temperature superconductivity in the absence of the magnetic-resonance mode. *Nature* **427**, 714–717 (2004).
- Zhu, J.-X. & Balatsky, A. V. Local strong coupling pairing in *d*-wave superconductor with inhomogeneous bosonic modes. Preprint at (<http://www.arXiv.org/cond-mat/0604391>) (2006).
- Pilgram, S., Rice, T. M. & Sigrist, M. Role of inelastic tunneling through the barrier in scanning tunneling microscope experiments on cuprates. Preprint at (<http://www.arXiv.org/cond-mat/0605646>) (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the following people for discussions and communications: Ar. Abanov, P. W. Anderson, N. Ashcroft, T. P. Devereaux, T. Egami, M. Eshrig, C. Henley, P. J. Hirschfeld, A. Lanzara, D.-H. Lee, P. Littlewood, D. Morr, K. A. Müller, M. R. Norman, J. Orenstein, T. M. Rice, J. Rowell, D. J. Scalapino, Z.-X. Shen, C. M. Varma and A. Zettl. This work was supported by an LDRD from Los Alamos National Laboratory, a Grant-in-Aid for Scientific Research from the Ministry of Science and Education (Japan), the 21st-Century COE Program of JSPS, by Cornell University and by the Office of Naval Research; K.F. acknowledges Fellowship support from the ICAM International Materials Institute.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.C.D. ([jcdavis@ccmr.cornell.edu](mailto:jcdavis@ccmr.cornell.edu)).



# Adaptive liquid microlenses activated by stimuli-responsive hydrogels

Liang Dong<sup>1</sup>, Abhishek K. Agarwal<sup>1</sup>, David J. Beebe<sup>2</sup> & Hongrui Jiang<sup>1</sup>

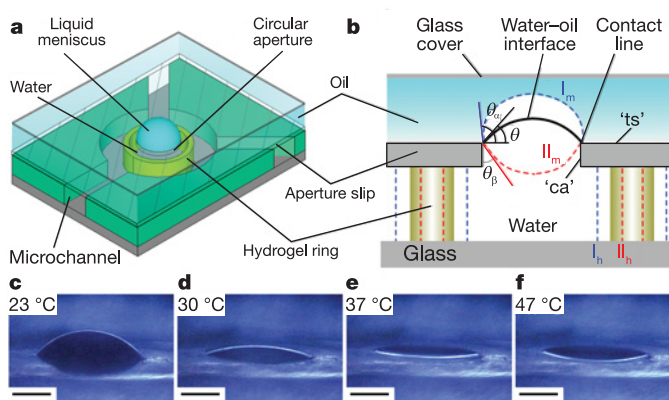
Despite its compactness, the human eye can easily focus on different distances by adjusting the shape of its lens with the help of ciliary muscles<sup>1</sup>. In contrast, traditional man-made optical systems achieve focusing by physical displacement of the lenses used. But in recent years, advances in miniaturization technology have led to optical systems that no longer require complicated mechanical systems to tune and adjust optical performance. These systems have found wide use in photonics, displays and biomedical systems. They are either based on arrays of microlenses with fixed focal lengths<sup>2–5</sup>, or use external control to adjust the microlens focal length<sup>6–12</sup>. An intriguing example is the tunable liquid lens, where electrowetting or external pressure manipulates the shape of a liquid droplet and thereby adjusts its optical properties. Here we demonstrate a liquid lens system that allows for autonomous focusing. The central component is a stimuli-responsive hydrogel<sup>13</sup> integrated into a microfluidic system and serving as the container for a liquid droplet, with the hydrogel simultaneously sensing the presence of stimuli and actuating adjustments to the shape—and hence focal length—of the droplet. By working at the micrometre scale where ionic diffusion and surface tension scale favourably<sup>14</sup>, we can use pinned liquid–liquid interfaces to obtain stable devices and realize response times of ten to a few tens of seconds. The microlenses, which can have a focal length ranging from  $-\infty$  to  $+\infty$  (divergent and convergent), are also readily integrated into arrays that may find use in applications such as sensing, medical diagnostics and lab-on-a-chip technologies<sup>15–19</sup>.

In our system, we use the meniscus between water and oil as an optical lens and adjust its focal length by changing the curvature of this meniscus (Supplementary Video 1). The basic design (Fig. 1) consists of a stimuli-responsive hydrogel ring placed within a microfluidic channel system, and sandwiched between a glass plate and an aperture slip, the latter with an opening centred over the ring. The microchannels are filled with water, and oil is placed on top of this structure and capped with a glass cover slip (Methods). The sidewall and bottom surface of the aperture ('ca' in Fig. 1b) are hydrophilic and the top surface ('ts' in Fig. 1b) is hydrophobic, which ensures that the water–oil meniscus is pinned along the hydrophobic–hydrophilic contact line 'ca–ts' (that is, the top edge of the aperture opening). When exposed to an appropriate stimulus (which could be pH<sup>20</sup>, temperature<sup>20</sup>, light<sup>21,22</sup>, an electric field<sup>23</sup>, antigens<sup>24</sup>, and so on), the hydrogel ring underneath the aperture opening responds by expanding or shrinking, owing to the absorption or release of water via the hydrogel network interstitials<sup>13</sup>; this leads to a change in the volume of the water droplet located in the middle of the ring. The net volume changes—the change in the volume enclosed by the ring, and the change in water droplet volume—cause a change in the pressure difference  $P$  across the water–oil interface, with  $P$  directly determining the geometry of the liquid meniscus. Because the contact line of the meniscus is pinned and stationary, volume

changes are translated into a change in curvature and hence angle  $\theta$  (Fig. 1b), which determines the focal length of the microlens. Angle  $\theta$  may attain any value in the interval  $-(90^\circ - \theta_\beta) \leq \theta \leq \theta_\alpha$  by varying  $P$  (Supplementary Information), where  $\theta_\alpha$  and  $\theta_\beta$  are the water contact angles on the 'ts' and 'ca' surfaces, respectively.

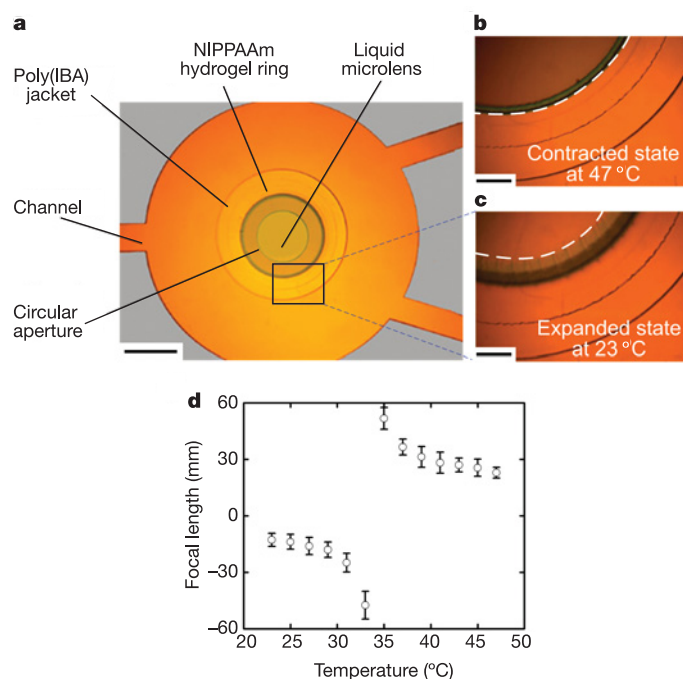
Here we use three hydrogels to produce the 'smart' attribute of the microlens, namely NIPAAm, a temperature-sensitive hydrogel, and AA and DMAEMA, which are pH-sensitive hydrogels (see Methods for details).

We first demonstrate the 'smart liquid microlens' concept with the temperature-sensitive NIPAAm hydrogel that expands at low temperatures and contracts at high temperatures, with a volume transition temperature of approximately 32 °C (Fig. 2 and Supplementary Information; see Methods for details on device fabrication). In this system, a polymer jacket placed around the hydrogel ring physically restricts the latter's movement so that it expands and contracts only at its inside periphery (Fig. 2b, c). At low temperatures, the liquid meniscus grows because the volume increase of the hydrogel upon expansion is larger than the volume of water lost from the ring due to absorption by the hydrogel; at high temperatures, the



**Figure 1 | Smart microlens using a pinned liquid–liquid interface.** **a**, The water–oil interface forms the liquid microlens. The microchannels allow the flow of fluids to the microlens structure. **b**, Smart variable-focus mechanism. The hydrophilic sidewall and bottom surface ('ca') and hydrophobic top surface ('ts') of the aperture pin the water–oil meniscus along the contact line 'ca–ts'. The expansion and contraction of the hydrogel regulates the shape of the liquid meniscus by changing the angle  $\theta$  of the pinned water–oil interface. The blue dashed lines show the expanded state of the hydrogel ring ('I<sub>h</sub>') and the corresponding divergent microlens ('I<sub>m</sub>') at  $\theta = \theta_\alpha$ . The red dashed lines show the contracted state of the hydrogel ring ('II<sub>h</sub>') and the corresponding convergent microlens ('II<sub>m</sub>') at  $\theta = -(90^\circ - \theta_\beta)$ . **c–f**, The shape of the liquid microlens varies with local environmental temperature. Scale bars, 1.0 mm.

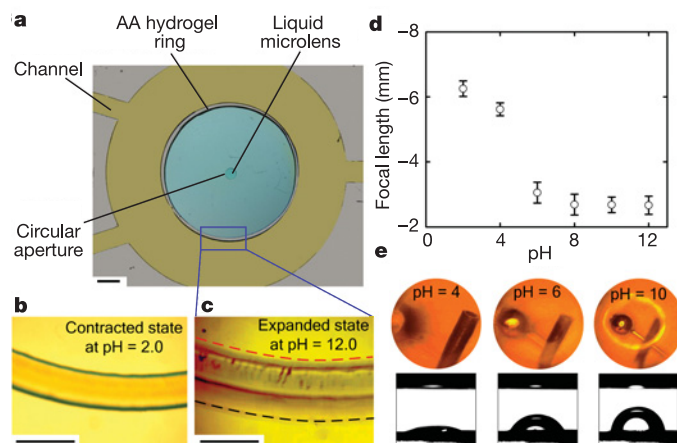
<sup>1</sup>Department of Electrical and Computer Engineering, <sup>2</sup>Department of Biomedical Engineering, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA.



**Figure 2 | A smart temperature-sensitive liquid microlens using NIPAAm hydrogel.** NIPAAm, *N*-isopropylacrylamide; IBA, isobornyl acrylate; see Methods for details. **a**, An optical image of the device. Scale bar, 2.0 mm. **b**, **c**, The contracted (**b**) and expanded (**c**) states of the hydrogel ring. The dashed lines show the boundaries of the inside periphery of the hydrogel ring at two temperature states. Scale bars, 500  $\mu$ m. **d**, The focal length of the microlens as a function of temperature. The microlens is divergent between 23 °C (focal length = -11.7 mm) and 33 °C (focal length approaches  $\infty$ ). Between 33 °C and 47 °C, the microlens becomes convergent with a positive focal length from  $+\infty$  to 22.8 mm. Error bars,  $\pm$ s.d.

liquid meniscus retreats because the water release accompanying hydrogel shrinkage is unable to compensate for the decrease in physical volume of the hydrogel ring (see Supplementary Fig. S1). The difference between the physical volume change of the hydrogel and the volume change in water outside the hydrogel is probably due to the different amount and structure of bound and free water within a swollen and collapsed hydrogel<sup>25</sup>. Figure 2d shows how the microlens adjusts its focal length from several millimetres to infinity in both positive and negative regimes as the temperature varies. For characterization only, an external heater is used here to change the local environmental temperature. A response time of 20–25 s is observed for the device; this refers to the time from turning on the heater to a visual change in the shape of the liquid meniscus (including 10–15 s of heater ramp and stabilization time).

Our second smart liquid microlens is fabricated using the pH-sensitive AA hydrogel, which expands in basic solutions and contracts in acidic solutions with a volume transition point at pH 5.5 (Fig. 3 and Supplementary Information; see Methods for device fabrication). In this system, the volume of the hydrogel ring is regulated by exposing its outside periphery to various pH buffers flowing through the microchannel (Fig. 3b, c). This effect ensures that as the buffer solution's pH value increases, the meniscus bulges upward and thereby decreases the focal length of the microlens (Fig. 3d; also see Supplementary Fig. S2). This observation indicates that the ring expansion and decrease in the volume enclosed inside the ring exceeds the decrease in water volume due to the absorption of the droplet by the surrounding hydrogel ring. The response time of the pH-sensitive microlens is approximately 12 s, as judged by exposing the hydrogel ring to the desired pH buffers and measuring the time interval until a visual change in the shape of the liquid meniscus is observable. Figure 3e illustrates the ability to focus on objects at

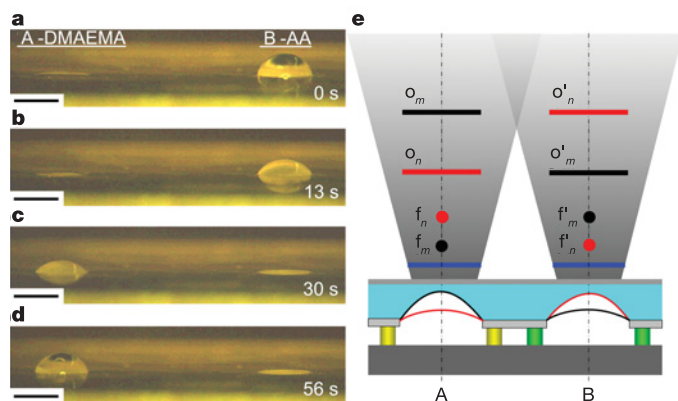


**Figure 3 | A smart pH-sensitive liquid microlens using AA hydrogel.** AA, acrylic acid; see Methods for details. **a**, An optical image of the device. Scale bar, 1.0 mm. **b**, The contracted state of the hydrogel ring at pH 2.0. Scale bar, 500  $\mu$ m. **c**, The expanded state of the hydrogel ring at pH 12.0, where the red and black dashed lines represent the boundaries of the inside and outside periphery of the hydrogel ring, respectively. Scale bar, 500  $\mu$ m. **d**, The focal length of the microlens as a function of pH. Error bars,  $\pm$ s.d. **e**, 'Smart' focusing of two objects (a ball connected to a pillar and a needle tip separated by 1.38 cm). At pH 2.0, both objects are out of focus (not shown). The top row optical images show (left to right): the needle tip is focused at pH 4.0; the focus point falls somewhere between the two objects at pH 6.0; and the ball is focused at pH 10.0. The bottom row shows the corresponding optical images of the liquid droplet itself.

different distances: the microlens device is focused on a needle tip when using a buffer solution with pH 4.0, and subsequently on a ball when changing the buffer pH to 10.0.

Inspired by insect (for example, *Drosophila*) compound eyes, where each repeating ommatidium<sup>26</sup> is responsible for one area in space and can progressively turn on and off as an object moves across the visual field of the eye, we extend the single liquid microlens to a microlens array. As a first step, we integrate two liquid microlenses in one microchannel to dynamically monitor two areas (Fig. 4 and Supplementary Information; see Methods for device fabrication). These two microlenses have identical structures, but are fabricated from two pH-sensitive hydrogels with opposite responses to pH changes: the AA hydrogel used already for the one-lens system expands at high pH and contracts at low pH, whereas DMAEMA hydrogel expands at low pH and contracts at high pH with a volume transition point of about pH 7.0. When replacing an initial high-pH buffer with a low-pH buffer, the liquid meniscus of the DMAEMA hydrogel-based microlens A gradually bulges up to move the focal plane closer to the aperture, while the AA hydrogel-based microlens B exhibits the opposite effect and moves the focal plane further away from the aperture (Fig. 4a–d and Supplementary Video 2). The object planes of microlenses A and B thus move through space perpendicular to the aperture as the focal lengths gradually change (Fig. 4e). In the array, each microlens provides information about one designated area. In this regard, the array mimics compound eyes where different unit eyes monitor different areas<sup>26</sup>; but the individual liquid microlenses within our array have in addition variable-focus ability and can in principle be designed to respond to different stimuli.

The present hydrogel-based microlens system is relatively simple and highly adaptable, particularly as appropriate modifications to the fabrication procedure (for example, controlling photo-definability, crosslinking density, and porosity) should make it possible to incorporate some of the many other hydrogels<sup>20–24</sup> that respond, for example, to light, electric fields and antigens. This should make it possible to incorporate microlenses controlled by different hydrogels into functionally complex arrays, which might then find use as physical, biological or chemical sensors that are capable of sensing



**Figure 4 | Combination of two smart pH-sensitive liquid microlenses to monitor two areas in space.** **a–d**, Optical images of the operation of the two-pixel pH-sensitive liquid microlens array, showing snapshots at indicated times. Left, microlens A, 2-(dimethylamino)ethyl methacrylate (DMAEMA); right, microlens B, AA. Scale bars, 500  $\mu\text{m}$ . **e**, Conceptual diagram showing the monitoring of different areas in space by lenses A and B. The blue lines represent the image planes. The red and black curves represent the oil–water interfaces of the two microlenses at two pH values,  $n$  and  $m$ , respectively. For microlens A, the focal point changes from  $f_n$  (red point) to  $f_m$  (black point) and the object plane moves from  $O_n$  (red line) to  $O_m$  (black line), as the pH value of its surrounding fluid changes from  $n$  to  $m$ . Microlens B exhibits opposite effects ( $f'_n$  to  $f'_m$  and  $O'_n$  to  $O'_m$ ) when simultaneously exposed to this same local environmental fluid.

multiple environmental parameters and generating optical outputs or visible images that are simple to read. The versatility of this system, together with the simple and direct sensing and read-out mechanisms, also mean that the hydrophobic–hydrophilic boundary interface at the ‘heart’ of the device can contain an aperture opening of an arbitrary shape (see Supplementary Fig. S3 for a cylindrical liquid microlens with a rectangular aperture). Moreover, the interface can be realized with a variety of substrates (for example, glass or polymers), with no strict geometrical substrate requirements. This should allow the fabrication of three-dimensional microlens arrays on flexible polymer substrates<sup>27</sup>, and thus the achievement of larger fields of view (see Supplementary Fig. S4 for a liquid microlens array on a curved surface).

A number of alternative microlens technologies allow for simple and cheap fabrication of small devices with fast response times (in the millisecond regime). The present hydrogel-based systems are much slower, but further miniaturization of the hydrogel structures<sup>28</sup> should improve the response times at least somewhat. Nonetheless, a unique feature of these systems is that they do not require complicated external control systems; instead, the microlenses can autonomously respond to changes intrinsic to a liquid sample being studied and provide visible output signals. Another attractive feature is that these microlenses can be readily integrated with existing electronic and opto-electronic systems (for example, simple planar electrodes can be used to precisely modulate the volumetric changes of an electrically responsive hydrogel<sup>29</sup>), with lens fabrication by *in situ* liquid-phase photopolymerization<sup>28</sup> facilitating the integration with existing microfluidic components<sup>30</sup>. It should therefore be not too difficult to produce functionally complex yet relatively simple systems—for example, diagnostic or lab-on-a-chip applications.

## METHODS

**Photopolymerizable mixtures.** The NIPAAm temperature-sensitive hydrogel prepolymer mixture consists of *N*-isopropylacrylamide (NIPAAm), *N,N'*-methylenebisacrylamide, dimethyl sulphoxide, deionized water and 2,2-dimethoxy-2-phenylacetophenone (DMPA) in the weight ratio of 2.18:0.124:3.0:1.0:0.154. The AA pH-sensitive hydrogel prepolymer mixture consists of acrylic acid (AA), 2-hydroxyethyl methacrylate (HEMA, 0–50 p.p.m. MEHQ inhibitor), ethylene glycol dimethacrylate (EGDMA) and DMPA in

the weight ratio of 4.054:29.286:0.334:1.0. The DMAEMA pH-sensitive hydrogel pre-polymer mixture consists of 2-(dimethylamino)ethyl methacrylate (DMAEMA), HEMA, EGDMA and DMPA in the weight ratio of 5.718:27.627:0.467:1.0. The microchannels and aperture slips of all devices, and the polymer jacket in Fig. 2, are made by photopatterning a non-responsive prepolymer mixture (called poly(IBA)) consisting of isobornyl acrylate (IBA), tetraethylene glycol dimethacrylate and DMPA in the weight ratio of 31.66:1.66:1.0.

**Fabrication of devices.** The devices in Figs 2–4 are fabricated by using liquid-phase photopolymerization (LP<sup>3</sup>) based on ultraviolet (UV) photolithography<sup>28</sup>. The fabrication starts with a 250- $\mu\text{m}$ -thick polycarbonate cartridge (HybriWells, Grace Bio-Labs) that is filled with the poly(IBA) prepolymer mixture. An aperture is formed inside the cartridge via direct photopatterning of the mixture through a photomask aligned on top of the cartridge (UV intensity  $I_{\text{UV,poly(IBA)}} = 7.7 \text{ mW cm}^{-2}$ ; exposure time  $t_{\text{UV,poly(IBA)}} = 24 \text{ s}$ ). The cartridge's thin liner is peeled off, leaving one side of the photopatterned poly(IBA) slip exposed to the ambient air. 100% ethanol is used to remove the unpolymerized pre-polymer. To make the poly(IBA) sidewall of the aperture hydrophilic, oxygen plasma treatment is carried out using a reactive ion etching system (Technics Micro-RIE series 800-IIC; power supply 70 W at 13.56 MHz; oxygen flow rate  $3.0 \text{ cm}^3 \text{ STP min}^{-1}$ ; gas pressure 107 mtorr; treatment time 40 s). Next, a cavity is formed by adhering the treated surface of the poly(IBA) slip to the glass substrate surface using double-sided adhesive spacer tapes along the edge of the slip. The cavity height is determined by the thickness of the adhesive tapes (750  $\mu\text{m}$  and 250  $\mu\text{m}$  for the microlenses in Fig. 2 and Figs 3–4, respectively). The poly(IBA) microchannels and jackets are constructed inside the cavity by using a sequential step-and-repeat LP<sup>3</sup> process ( $I_{\text{UV,poly(IBA)}} = 8.5 \text{ mW cm}^{-2}$  and  $t_{\text{UV,poly(IBA)}} = 34.5 \text{ s}$ ). The polycarbonate cartridge cover is finally peeled off of the poly(IBA) aperture slip, which now serves as the top slip of the microchannels.

The NIPAAm hydrogel ring (Fig. 2) and AA hydrogel ring (Figs 3–4) are constructed by photopatterning the corresponding aforementioned hydrogel liquid prepolymer mixtures. The photolithography conditions are:  $I_{\text{UV,NIPAAm}} = 12.5 \text{ mW cm}^{-2}$  and  $t_{\text{UV,NIPAAm}} = 8.5 \text{ s}$ ;  $I_{\text{UV,AA}} = 10 \text{ mW cm}^{-2}$  and  $t_{\text{UV,AA}} = 45.3 \text{ s}$ . Next, the DMAEMA hydrogel ring (Fig. 4), which has an identical structure to its neighbouring AA hydrogel ring, is fabricated in the same microchannel as the AA hydrogel ring using the sequential step-and-repeat LP<sup>3</sup> process ( $I_{\text{UV,DMAEMA}} = 19.0 \text{ mW cm}^{-2}$  and  $t_{\text{UV,DMAEMA}} = 50 \text{ s}$ ). The top-side surface of the aperture slip is rendered hydrophobic by manually coating it with an octadecyltrichlorosilane solution diluted by hexadecane (0.2 vol.%). A liquid meniscus is initially formed by manually loading water into the hydrogel ring through the aperture. The initial curvature of the meniscus depends on the amount of the loaded water. To store the oil on top of the water, a pre-fabricated fence made of polydimethylsiloxane elastomer is adhered to the aperture slip. The fence is then capped by a 100- $\mu\text{m}$ -thick glass cover slip.

Received 23 December 2005; accepted 23 June 2006.

- Toates, F. M. Accommodation function of human eye. *Physiol. Rev.* **52**, 828–863 (1972).
- Popovic, Z. D., Sprague, R. A. & Connell, G. A. N. Technique for monolithic fabrication of microlens arrays. *Appl. Opt.* **27**, 1281–1284 (1988).
- Cayre, O. J. & Pounov, V. N. Fabrication of microlens arrays by gel trapping of self-assembled particle monolayers at the decane–water interface. *J. Mater. Chem.* **14**, 3300–3302 (2004).
- Yang, R., Wang, W. J. & Soper, S. A. Out-of-plane microlens array fabricated using ultraviolet lithography. *Appl. Phys. Lett.* **86**, 161110 (2005).
- Yang, S. *et al.* Functional biomimetic microlens arrays with integrated pores. *Adv. Mater.* **17**, 435–438 (2005).
- Zhang, D. Y., Lien, V., Berdichevsky, Y., Choi, J. & Lo, Y. H. Fluidic adaptive lens with high focal length tunability. *Appl. Phys. Lett.* **82**, 3171–3172 (2003).
- Saitoh, A. & Tanaka, K. Self-developing aspherical chalcogenide-glass microlenses for semiconductor lasers. *Appl. Phys. Lett.* **83**, 1725–1727 (2003).
- Hayes, R. A. & Feenstra, B. J. Video-speed electronic paper based on electrowetting. *Nature* **425**, 383–385 (2003).
- Kuiper, S. & Hendriks, B. H. W. Variable-focus liquid lens for miniature cameras. *Appl. Phys. Lett.* **85**, 1128–1130 (2004).
- Ren, H. W., Fan, Y. H. & Wu, S. T. Liquid-crystal microlens arrays using patterned polymer networks. *Opt. Lett.* **29**, 1608–1610 (2004).
- Lopez, C. A., Lee, C. C. & Hirs, A. H. Electrochemically activated adaptive liquid lens. *Appl. Phys. Lett.* **87**, 134102 (2005).
- Ren, H. & Wu, S. T. Variable-focus liquid lens by changing aperture. *Appl. Phys. Lett.* **86**, 211107 (2005).
- Osada, Y., Gong, J. P. & Tanaka, Y. Polymer gels. *J. Macromol. Sci. C* **44**, 87–112 (2004).
- Atencia, J. & Beebe, D. J. Controlled microfluidic interfaces. *Nature* **437**, 648–655 (2005).



15. Camou, S., Fujita, H. & Fujii, T. PDMS 2D optical lens integrated with microfluidic channels: Principle and characterization. *Lab Chip* **3**, 40–45 (2003).
16. Chronis, N., Liu, G. L., Jeong, K. H. & Lee, L. P. Tunable liquid-filled microlens array integrated with microfluidic network. *Opt. Express* **11**, 2370–2378 (2003).
17. Burns, M. A. *et al.* An integrated nanoliter DNA analysis device. *Science* **282**, 484–487 (1998).
18. Choi, H. W. *et al.* GaN micro-light-emitting diode arrays with monolithically integrated sapphire microlenses. *Appl. Phys. Lett.* **84**, 2253–2255 (2004).
19. Carlson, K. *et al.* *In vivo* fiber-optic confocal reflectance microscope with an injection-molded plastic miniature objective lens. *Appl. Opt.* **44**, 1792–1797 (2005).
20. Park, T. G. & Hoffman, A. S. Synthesis and characterization of pH- and temperature-sensitive hydrogels. *J. Appl. Polym. Sci.* **46**, 659–671 (1992).
21. Suzuki, A. & Tanaka, T. Phase transition in polymer gels induced by visible light. *Nature* **346**, 345–347 (1990).
22. Sershen, S. R. *et al.* Independent optical control of microfluidic valves formed from optomechanically responsive nanocomposite hydrogels. *Adv. Mater.* **17**, 1366–1368 (2005).
23. Tanaka, T., Nishio, I., Sun, S. T. & Uenonishio, S. Collapse of gels in an electric field. *Science* **218**, 467–469 (1982).
24. Miyata, T., Asami, N. & Uragami, T. A reversibly antigen-responsive hydrogel. *Nature* **399**, 766–769 (1999).
25. Lele, A. K., Hirve, M. M., Badiger, M. V. & Mashelkar, R. A. Predictions of bound water content in poly(*N*-isopropylacrylamide) gel. *Macromolecules* **30**, 157–159 (1997).
26. Land, M. F. Visual acuity in insects. *Annu. Rev. Entomol.* **42**, 147–177 (1997).
27. Jacobs, H. O., Tao, A. R., Schwartz, A., Gracias, D. H. & Whitesides, G. M. Fabrication of a cylindrical display by patterned assembly. *Science* **296**, 323–325 (2002).
28. Beebe, D. J. *et al.* Functional hydrogel structures for autonomous flow control inside microfluidic channels. *Nature* **404**, 588–590 (2000).
29. Bassetti, M. J., Chatterjee, A. N., Aluru, N. R. & Beebe, D. J. Development and modeling of electrically triggered hydrogels for microfluidic applications. *J. Microelectromech. Syst.* **14**, 1198–1207 (2005).
30. Agarwal, A. K., Sridharamurthy, S. S., Beebe, D. J. & Jiang, H. Programmable autonomous micromixers and micropumps. *J. Microelectromech. Syst.* **14**, 1409–1421 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This research was supported in part by the US Department of Homeland Security, through a grant awarded to the National Center for Food Protection and Defense at the University of Minnesota, and in part by the Wisconsin Alumni Research Foundation (WARF). The authors thank J. Moorthy and S.S. Sridharamurthy for discussions.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.J. ([hongrui@engr.wisc.edu](mailto:hongrui@engr.wisc.edu)).

# Plant litter decomposition in a semi-arid ecosystem controlled by photodegradation

Amy T. Austin<sup>1</sup> & Lucía Vivanco<sup>1</sup>

The carbon balance in terrestrial ecosystems is determined by the difference between inputs from primary production and the return of carbon to the atmosphere through decomposition of organic matter<sup>1</sup>. Our understanding of the factors that control carbon turnover in water-limited ecosystems is limited, however, as studies of litter decomposition have shown contradictory results and only a modest correlation with precipitation<sup>2–5</sup>. Here we evaluate the influence of solar radiation, soil biotic activity and soil resource availability on litter decomposition in the semi-arid Patagonian steppe using the results of manipulative experiments carried out under ambient conditions of rainfall and temperature. We show that intercepted solar radiation was the only factor that had a significant effect on the decomposition of organic matter, with attenuation of ultraviolet-B and total radiation causing a 33 and 60 per cent reduction in decomposition, respectively. We conclude that photodegradation is a dominant control on above-ground litter decomposition in this semi-arid ecosystem. Losses through photochemical mineralization may represent a short-circuit in the carbon cycle, with a substantial fraction of carbon fixed in plant biomass being lost directly to the atmosphere without cycling through soil organic matter pools. Furthermore, future changes in radiation interception due to decreased cloudiness, increased stratospheric ozone depletion, or reduced vegetative cover may have a more significant effect on the carbon balance in these water-limited ecosystems than changes in temperature or precipitation.

Traditional models of the controls on litter decomposition in terrestrial ecosystems have focused on how soil organisms interact with climate and litter quality to control mass loss and nutrient release of senescent plant material<sup>6–8</sup>. Although litter decomposition does positively correlate with annual precipitation at the regional scale<sup>6,9,10</sup>, these empirical models cannot account for the rapid turnover of organic material observed in arid ecosystems<sup>2,3,11</sup>, or the lack of correlation of decomposition with rainfall in deserts<sup>4,5</sup>. These conflicting results suggest that there are unique factors affecting decomposition in arid and semi-arid ecosystems, which may include abiotic controls such as photodegradation or limitation on biotic activity from low soil resource availability. At the same time, the net effect on carbon sequestration in terrestrial ecosystems will result from the balance between changes in primary production and decomposition<sup>12,13</sup>, and as such, the controls on decomposition in water-limited ecosystems are crucial parameters for predicting future impacts of global change.

Photochemical mineralization of organic material results in a reduction of the average molecular mass of organic compounds, alteration of the capacity to absorb light both in the ultraviolet and visible spectrum, and the formation of novel photoproducts<sup>14,15</sup>. Solar radiation plays an important role in the turnover of organic

matter in aquatic ecosystems and oceans: photochemical reactions change the quality of dissolved organic matter (DOM)<sup>14,16</sup> and produce dissolved inorganic carbon and volatile CO<sub>2</sub>, CO and carbonyl sulphides<sup>15,17</sup>. Photochemical production of pyruvate from DOM in ocean surface waters, for example, is believed to be a critical transformation to permit biological degradation of recalcitrant DOM<sup>14</sup>. Although dissolved inorganic carbon (in the form of CO) is quantitatively one of the most important photoproducts observed during photochemical mineralization of DOM<sup>16</sup>, most studies suggest that the indirect effects on lability of organic matter, and not direct production of photoproducts, are the most important influences of solar radiation on the carbon cycle in aquatic ecosystems<sup>14,16</sup>.

There are very few studies of the direct effects of solar radiation on carbon turnover in terrestrial ecosystems, although CO<sub>2</sub> production from sterilized litter subjected to radiation treatments has been observed in emergent macrophytic vegetation<sup>18</sup> and carbon monoxide has been detected as a photoproduct from live and senescent plant material in short-term incubations in grasslands<sup>17,19</sup>. Ultraviolet-B radiation (UV-B) has been shown to affect litter decomposition through changes in the chemical composition of the litter, or through changes in the microbial community characteristics<sup>20,21</sup>. However, large direct UV-B effects on carbon turnover are undocumented, and only small negative and positive direct effects have been observed<sup>21–23</sup>. In contrast, simulations for deserts using the ecosystem model CENTURY suggest that direct effects of UV-B through increased photodegradation of litter could be important<sup>24</sup>. Photodegradation has been mentioned as a possible control on carbon turnover in studies of litter decomposition in water-limited ecosystems<sup>2,11</sup>, although this effect has never been quantified empirically. As such, the quantitative importance at the ecosystem scale of photodegradation effects on carbon turnover in arid and semi-arid ecosystems, from both direct effects on the production of volatile compounds, and indirect effects on the quality of organic matter, is unknown.

We conducted two factorial experiments under ambient conditions of rainfall and temperature designed to disentangle the abiotic and biotic controls on litter decomposition in a semi-arid ecosystem. In the first experiment, recently senesced litter of mixed native grasses was decomposed under different light regimes (full sunlight, reduced UV-B radiation, and blocked total radiation) in combination with a treatment of soil biocide to reduce biotic activity. In a second parallel experiment, the same litter mixture was decomposed, treating underlying soils with additions of labile carbon and inorganic nitrogen to stimulate biotic activity and remove possible resource limitation for decomposer organisms.

Manipulations of biotic activity and resource availability produced dramatic changes in soil characteristics. There was a significant decline in the soil microbial biomass C ( $P < 0.05$ ), potential soil respiration ( $P < 0.001$ ) and soil nitrate concentrations ( $P < 0.05$ ) of

<sup>1</sup>Instituto de Investigaciones Fisiológicas y Ecológicas Vinculadas a la Agricultura (IFEVA) and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Facultad de Agronomía, Universidad de Buenos Aires, Av. San Martín 4453, Buenos Aires (C1417DSE), Argentina.

**Table 1 | Soil characteristics of litter decomposition experiments**

|                       | C mineralization<br>( $\mu\text{g C per g soil d}^{-1}$ ) | Microbial biomass<br>( $\mu\text{g C per g dry soil}$ ) | Soil $\text{NH}_4\text{-N}$<br>( $\mu\text{g per g dry soil}$ ) | Soil $\text{NO}_3\text{-N}$<br>( $\mu\text{g per g dry soil}$ ) | Soil $\text{H}_2\text{O}$<br>(%) <sup>*</sup> |
|-----------------------|-----------------------------------------------------------|---------------------------------------------------------|-----------------------------------------------------------------|-----------------------------------------------------------------|-----------------------------------------------|
| Radiation and biocide |                                                           |                                                         |                                                                 |                                                                 |                                               |
| (-)Biocide            | $5.3 \pm 0.50^a$                                          | $121 \pm 15^a$                                          | $3.31 \pm 0.35^a$                                               | $0.64 \pm 0.09^a$                                               | $2.5 \pm 0.21^a$                              |
| (+)Biocide            | $1.5 \pm 0.87^b$                                          | $64 \pm 18^b$                                           | $9.26 \pm 1.40^b$                                               | $0.20 \pm 0.08^b$                                               | $2.8 \pm 0.25^a$                              |
| Substrate addition    |                                                           |                                                         |                                                                 |                                                                 |                                               |
| Control               | $6.1 \pm 0.60^a$                                          | $111 \pm 20^a$                                          | $3.99 \pm 0.66^a$                                               | $0.76 \pm 0.09^a$                                               | $3.7 \pm 0.24^a$                              |
| Carbon                | $18.1 \pm 2.47^b$                                         | $223 \pm 40^b$                                          | $0.86 \pm 0.36^a$                                               | $0.03 \pm 0.01^a$                                               | $2.8 \pm 0.28^{ab}$                           |
| Nitrogen              | $6.4 \pm 0.68^a$                                          | $96 \pm 19^a$                                           | $42.39 \pm 4.99^b$                                              | $3.97 \pm 0.70^b$                                               | $2.5 \pm 0.08^b$                              |
| C and N               | $13.6 \pm 0.57^b$                                         | $386 \pm 74^b$                                          | $7.38 \pm 5.35^a$                                               | $0.58 \pm 0.37^a$                                               | $2.7 \pm 0.31^b$                              |

Experiments were performed in the Patagonian steppe; data shown were obtained at the last litter harvest. Data shown for potential C mineralization are laboratory incubations of 7 days ( $n = 5$ ) and microbial biomass C, soil inorganic N and soil water from field measurements ( $n = 5$ ). Mean values are shown ( $\pm$ s.e.m.). Different letters indicate significant differences among treatments for Tukey post-hoc comparisons ( $P < 0.05$ ).

<sup>\*</sup> Calculated as g per g dry soil.

biocide-treated soils (Table 1). In addition, there was a significant effect of biocide application on microbial colonization of litter for both cultivable fungal and bacterial populations (Table 2,  $P < 0.001$ ). The biocide application eliminated cultivable litter fungal populations entirely and reduced cultivable litter bacterial colonization by an order of magnitude. In the substrate addition experiment, soil additions of labile C and inorganic N resulted in increased soil microbial biomass ( $P < 0.05$ ), increased  $\text{NO}_3^-$  and  $\text{NH}_4^+$  concentrations in the N addition treatment ( $P < 0.0001$ ), and increased potential C mineralization with both C and C + N additions (Table 1,  $P < 0.01$ ).

In spite of considerable changes in biotic activity and soil resource availability, solar radiation was the only variable that significantly affected litter decomposition (Fig. 1,  $P < 0.001$ ). Decomposition, expressed as a constant ( $k$ ) that integrates the rates of organic matter loss over time, showed a 33% and 60% reduction in litter decomposition in the reduced UV-B and blocked total radiation treatments, respectively (Fig. 1b,  $P < 0.0001$ ). In contrast, there was no effect of biocide application on rates of organic matter loss or  $k$  constants of decomposition (Fig. 1,  $P > 0.05$ ); moreover, soil substrate additions had no effect on litter decomposition (Fig. 2,  $P > 0.05$ ). Finally, there was no interaction between radiation and biocide treatments for organic matter loss ( $P > 0.05$ ), suggesting that the principal effect of radiation exclusion was due to direct effects on photochemical mineralization of organic matter and not interaction with increased lability of substrates for soil biota. The organic matter loss observed in the total blocked radiation treatments (10% at the end of the incubation period, Fig. 1a) could be the result of abiotic processes other than photodegradation, such as physical fragmentation or leaching. At the same time, these possible effects do not appear to have been biased by soil temperature differences in the radiation

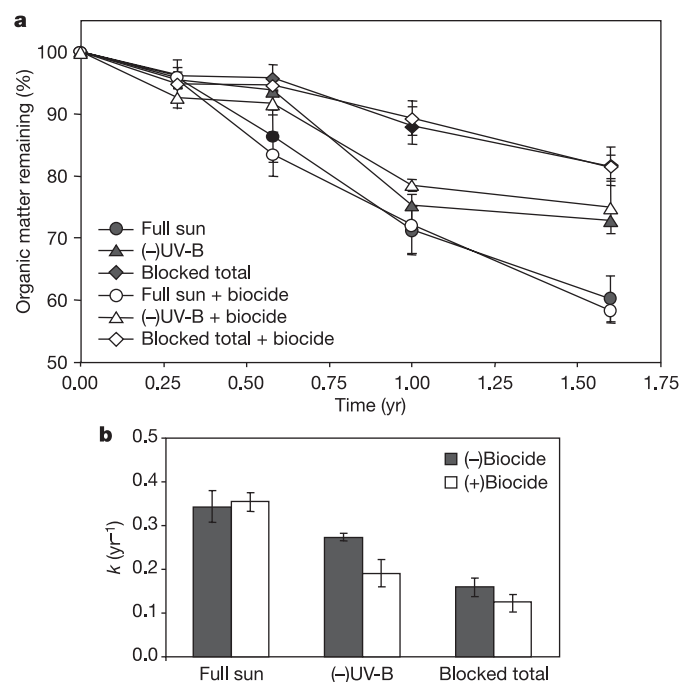
treatments. Overall mean temperature among radiation treatments from all sampling dates did not differ significantly ( $P > 0.05$ , see Supplementary Information).

Taken together, these results suggest that photodegradation is a dominant control of above-ground litter decomposition in this semi-arid ecosystem, and that UV-B radiation alone can account for almost 50% of the carbon lost due to photochemical mineralization of litter, a much larger effect of UV-B than has been previously reported<sup>22–25</sup>. Ultraviolet-A (UV-A) and/or short wavelength visible radiation have been shown to affect gaseous carbon losses from terrestrial vegetation<sup>18,19</sup>, and the results from our study support the observation that solar radiation other than UV-B also affected photodegradation, seen in the significant differences in organic matter loss between the attenuated UV-B and total blocked radiation treatments. Surprisingly, alteration of biotic activity, through inhibition with biocide and stimulation with labile carbon and nitrogen

**Table 2 | Effect of biocide and substrate addition on microbial colonization of litter**

| Experiment         | Treatment/substrate | Fungi (CFU $\text{d}^{-1}$ ) | Bacteria (CFU $\text{d}^{-1}$ ) |
|--------------------|---------------------|------------------------------|---------------------------------|
| (-)Biocide         | Full sun            | $9,050 \pm 3,460$            | $195 \pm 92$                    |
| (-)Biocide         | (-)UV-B             | $3,037 \pm 545$              | $153 \pm 66$                    |
| (-)Biocide         | Blocked total       | $22,470 \pm 3,865$           | $361 \pm 104$                   |
| (+)Biocide         | Full sun            | 0                            | $3 \pm 1$                       |
| (+)Biocide         | (-)UV-B             | 0                            | $20 \pm 9$                      |
| (+)Biocide         | Blocked total       | 0                            | $22 \pm 10$                     |
| Substrate addition | Control             | $1,125 \pm 982$              | $245 \pm 56$                    |
| Substrate addition | C                   | $1,235 \pm 745$              | $162 \pm 23$                    |
| Substrate addition | N                   | $604 \pm 150$                | $325 \pm 84$                    |
| Substrate addition | C and N             | $770 \pm 438$                | $301 \pm 176$                   |

Mean values ( $n = 5$ ) are shown ( $\pm$ s.e.m.) for colony forming units (CFUs) of fungi and bacteria from litter extracts after 24 h of cultivation. Litter from substrate addition experiments was not subject to either biocide or radiation treatments.



**Figure 1 | Effect of solar radiation and biocide on litter decomposition in the Patagonian steppe. a**, Organic matter remaining over time. Symbols indicate different treatments of radiation exclusion and biocide application ( $n = 5$ ); see Methods for details. Error bars,  $\pm$ s.e.m. **b**, Constants of decomposition,  $k$ . These were obtained by plotting  $\ln(\text{organic matter remaining}/\text{initial organic matter})$  against time;  $k$  is the slope of the regression ( $n = 5$ ). Error bars,  $\pm$ s.e.m.



additions, had no effect on litter decomposition. These results suggest that biotic activity exerts very little influence over rates of organic matter loss of above-ground litter in this ecosystem. Moreover, indirect effects of photodegradation resulting in changes in the lability of organic compounds, although it cannot be discounted entirely, appear to be secondary to direct photochemical mineralization of organic matter, a contrast with the relative importance of the two effects observed in studies of aquatic ecosystems<sup>15,16</sup>.

The importance of photodegradation for organic matter turnover in the Patagonian steppe suggests a 'short-circuit' in the carbon cycle, with implications for the functioning of semi-arid ecosystems. The loss of carbon through inorganic photoproducts may be a major pathway of carbon loss from above-ground plant biomass in this ecosystem, with a direct return of inorganic carbon to the atmosphere without cycling through soil organic matter pools. Patchy ecosystem structure with over 50% exposed bare soil<sup>26,27</sup>, high radiation interception at the soil surface, large amounts of standing dead material and a low number of rainy days are typical of arid and semi-arid ecosystems. These characteristics combine a set of unique conditions under which photodegradation can dominate decomposition of above-ground litter, and could explain the direct, large effects of UV-B radiation that have not been observed in more humid ecosystems at higher latitudes with contrasting dominant vegetation and litter quality<sup>21–23</sup>. A quantification at the ecosystem scale of this 'short-circuit' in the carbon cycle could rectify current discrepancies in modelled carbon balances of arid and semi-arid ecosystems, and explain the lack of correlation with traditional models of biotic controls on decomposition based on variation in climate and litter quality<sup>6</sup>.

Ecosystem-scale processes governing the carbon balance in water-limited ecosystems may respond differently to global change owing to differential controls on production and decomposition<sup>9</sup>. Effects of global change on carbon fixation in arid and semi-arid ecosystems include direct effects of elevated carbon dioxide on net primary production<sup>28</sup>, as well as indirect effects on shrub/grass ratio, vegetative cover and climate patterns<sup>29,30</sup>. Our results suggest that the direct

effects of solar radiation on above-ground decomposition overshadow the importance of other controls such as water availability, which have shown little or no correlation with litter decomposition in deserts<sup>3,4</sup>, and smaller effects on organic mass loss in rainfall manipulation experiments<sup>5,27</sup>. Changes in radiation dose due to decreased cloud cover, increased ozone depletion or reduced vegetative cover could directly affect carbon loss<sup>19</sup>, and could be more important than changes in rainfall amount or climatic variability. As close to 40% of the terrestrial land surface is currently classified as arid or semi-arid, our understanding of how human-induced global change will affect key controls on the carbon cycle in these ecosystems is critical, and factors affecting rates of photochemical mineralization could have consequences for the potential of carbon sequestration in water-limited ecosystems and the global carbon balance.

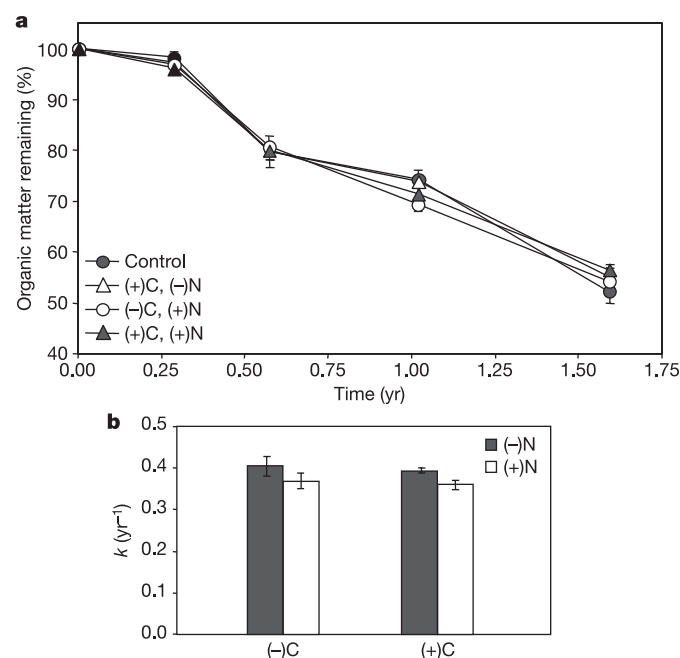
## METHODS

**Study site.** The Instituto Nacional de Tecnología Agropecuaria (INTA) Río Mayo experimental station is located in the Argentinean province of Chubut (45° 41' S, 70° 16' W) at 500 m elevation. Long-term mean annual precipitation is 152 mm, and is strongly seasonal, with >70% of the precipitation falling in winter months. Monthly mean temperature ranges from 15 °C in January to 1 °C in July. The vegetation is classified as semi-arid steppe, with the dominant vegetation of perennial tussock grasses (32% basal cover) and shrubs (15% basal cover)<sup>26</sup>. Soils are coarsely textured aridisols, with high gravel content and low soil-water holding capacity. Solar irradiance for the period of study is shown in Supplementary Fig. 2.

**Field methods.** For both experiments, a representative mixture of leaf litter from the site's dominant perennial grass species (*Stipa speciosa*, *Poa ligularis*, *Stipa humilis*) was collected and separated from green and standing dead material to include only recently-senesced litter. We established 50 plots of 1 m<sup>2</sup> within a grazing enclosure. In each plot, we removed all above-ground vegetation to minimize shading and plant-microbial competition. For the radiation/biocide experiment, plastic frames of 20 × 10 × 6 cm were used to construct litter boxes. We attached a plastic filter on the top and the northern face of the frame that corresponded to one of the three light treatments: (1) Aclar filters, which allowed transmission of >95% of solar radiation (full sun); (2) Mylar filters, which attenuated all radiation below 310 nm, '(–)UV-B'; and (3) Mylar filters covered with reflective aerosol paint that effectively blocked >90% of solar radiation (blocked total). Perforations were made in the filters to allow water infiltration by precipitation. Fibreglass mesh (2 mm) was attached to the underside and lateral sides of the boxes, so that the litter was in direct contact with the soil but received adequate ventilation. We placed one gram of recently senesced litter in five litterboxes in each randomly assigned treatment plot, which was representative of the spatial distribution of standing dead and litter detritus exposed to solar radiation in this ecosystem. The soil and litter biocide treatment combined a general fungicide at 17.5 g m<sup>–2</sup> (Captan, containing N-(trichloromethylthio) phthalimide) with a bactericide at 15.0 g m<sup>–2</sup> (Agri-micina, containing streptomycin sulphate and oxytetracycline) in an aqueous suspension distributed uniformly (21 m<sup>–2</sup>) over the plot area. In addition, naphthalene pellets (100 g m<sup>–2</sup>) were placed on the soil surface and incorporated into the surface soil by raking. In control plots, water alone was applied at 21 m<sup>–2</sup>. Applications of biocide or water were made three times per year over the experimental period. At each sampling date, filters were evaluated for damage or age, and replaced as necessary. Soil temperature under litterboxes was evaluated using thermistors buried at 2 cm depth at four time points during the day.

For the substrate addition experiment, litterbags (20 × 10 cm) of fibreglass mesh (2 mm) filled with one gram of recently-senesced litter were placed ( $n = 5$ ) on the soil surface. Before litter placement and three times per year over the experimental period, substrate additions of labile carbon, a combination of glucose and cornstarch (annual rate of 330 g C m<sup>–2</sup> yr<sup>–1</sup>), and inorganic nitrogen, ammonium nitrate (annual rate of 40 kg N ha<sup>–1</sup> yr<sup>–1</sup>), were incorporated into the first 5 cm of the soil with gentle raking. Litterboxes and litterbags were collected at 3.5, 7, 12 and 18 months and analysed for changes in organic matter loss over time. Ash-free dry mass was determined for all samples to correct for soil contamination from the field.

**Characteristics and rates of soil biotic activity.** Gravimetric soil water content and inorganic N content were determined for all dates (all data not shown); inorganic N was evaluated using 2 N KCl extracts and measured colorimetrically with an Alpchem autoanalyser. Extractions of soil microbial biomass, potential soil C mineralization and bacterial and fungal colonization of litter for biocide and substrate addition treatments were completed at the last litter harvest. Microbial C was measured using a modified chloroform fumigation-extraction



**Figure 2 | Effect of soil substrate additions of labile carbon and nitrogen on litter decomposition in the Patagonian steppe.** **a**, Organic matter remaining over time. Symbols indicate different treatments of soil substrate additions ( $n = 5$ ); see Methods for details. Error bars,  $\pm$ s.e.m. **b**, Constants of decomposition,  $k$  ( $n = 5$ ). Error bars,  $\pm$ s.e.m.

technique with a conversion factor of 0.45 ( $K_c$ ). Potential soil carbon mineralization was determined in laboratory incubations at 25 °C for 7 days using 50 g of fresh soil, with addition of base traps of 0.1 M NaOH, titrated with 0.15 M BaCl<sub>2</sub> and 0.1 M HCl. Bacterial and fungal counts of litter were completed with subsamples of 50–100 mg of litter from the final litter collection which were extracted in a buffer solution (0.88% NaCl); 5 ml of suspension were spread-plated on nutritive agar for determination of bacteria (nutritive agar) and fungi (potato-dextrose agar, with 50 µg ml<sup>-1</sup> tetracycline and 200 µg ml<sup>-1</sup> streptomycin). Plates were incubated at 25 °C for 24 h and colony-forming units (CFUs) were counted.

**Statistics.** Data for the two experiments were analysed separately; a two-factor analysis of variance (ANOVA) was completed for evaluation of litter decomposition at each sampling date, for fungal/bacterial colonization, and  $k$  constants. Soil characteristics were evaluated using a one- or two-way ANOVA. When necessary, data were log-transformed to correct for violations of homogeneity of variance.

Received 16 January; accepted 28 June 2006.

- Olson, J. S. Energy storage and the balance of producers and decomposers in ecological systems. *Ecology* **44**, 322–331 (1963).
- Montaña, C., Ezcurra, E., Carrillo, A. & Delhoume, J. P. The decomposition of litter in grasslands of northern Mexico: a comparison between arid and non-arid environments. *J. Arid Environ.* **14**, 55–60 (1988).
- Moorhead, D. L. & Reynolds, J. F. Mechanisms of surface litter mass loss in the northern Chihuahuan desert: a reinterpretation. *J. Arid Environ.* **16**, 157–163 (1989).
- Steinberger, Y., Shmida, A. & Whitford, W. G. Decomposition along a rainfall gradient in the Judean desert, Israel. *Oecologia* **82**, 322–324 (1990).
- Whitford, W. G. *et al.* Rainfall and decomposition in the Chihuahuan desert. *Oecologia* **68**, 512–515 (1986).
- Meentemeyer, V. Macroclimate and lignin control of litter decomposition rates. *Ecology* **59**, 465–472 (1978).
- Swift, M. J., Heal, O. W. & Anderson, J. M. *Decomposition in Terrestrial Ecosystems* (Univ. California Press, Berkeley, 1979).
- Aber, J. D. & Melillo, J. M. Nitrogen immobilization in decaying hardwood leaf litter as a function of initial nitrogen and lignin content. *Can. J. Bot.* **60**, 2263–2269 (1982).
- Austin, A. T. Differential effects of precipitation on production and decomposition along a rainfall gradient in Hawai'i. *Ecology* **83**, 328–338 (2002).
- Gholz, H. L., Wedin, D. A., Smitherman, S. M., Harmon, M. E. & Parton, W. J. Long-term dynamics of pine and hardwood litter in contrasting environments: toward a global model of decomposition. *Glob. Change Biol.* **6**, 751–765 (2000).
- Vossbrinck, C. R., Coleman, D. C. & Wooley, T. A. Abiotic and biotic factors in litter decomposition in a semiarid grassland. *Ecology* **60**, 265–271 (1979).
- Schlesinger, W. H. & Lichter, J. Limited carbon storage in soil and litter of experimental forest plots under increased atmospheric CO<sub>2</sub>. *Nature* **411**, 466–469 (2001).
- Nemani, R. R. *et al.* Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* **300**, 1560–1563 (2003).
- Kieber, D. J., McDaniel, J. & Mopper, K. Photochemical source of biological substrates in sea water: implications for carbon cycling. *Nature* **341**, 637–639 (1989).
- Miller, W. L. & Zepp, R. G. Photochemical production of dissolved inorganic carbon from terrestrial organic matter: significance to the oceanic organic carbon cycle. *Geophys. Res. Lett.* **22**, 417–420 (1995).
- Mopper, K. *et al.* Photochemical degradation of dissolved organic carbon and its impact on the oceanic carbon cycle. *Nature* **353**, 60–62 (1991).
- Tarr, M. A., Miller, W. L. & Zepp, R. G. Direct carbon monoxide production from plant matter. *J. Geophys. Res.* **100**, 11403–11413 (1995).
- Anesio, A. M., Tranvik, L. J. & Granéli, W. Production of inorganic carbon from aquatic macrophytes by solar radiation. *Ecology* **80**, 1852–1859 (1999).
- Schade, G. W., Hormann, R. M. & Crutzen, P. J. CO emissions from degrading plant matter. *Tellus B* **51**, 899–908 (1999).
- Johnson, D., Campbell, C. D., Lee, J. A. & Callaghan, T. V. Arctic microorganisms respond more to elevated UV-B radiation than CO<sub>2</sub>. *Nature* **416**, 82–83 (2002).
- Pancotto, V. A., Sala, O. E., Robson, T. M., Caldwell, M. M. & Scopel, A. L. Direct and indirect effects of solar ultraviolet-B radiation on long-term decomposition. *Glob. Change Biol.* **11**, 1982–1989 (2005).
- Gehrke, C., Johanson, U., Callaghan, T. V., Chadwick, D. & Robinson, C. H. The impact of enhanced UV-B radiation on litter quality and decomposition processes in *Vaccinium* leaves from the subarctic. *Oikos* **72**, 213–222 (1995).
- Moody, S. A. *et al.* The direct effects of UV-B radiation on *Betula pubescens* litter decomposing at four European field sites. *Plant Ecol.* **154**, 29–36 (2001).
- Moorhead, D. L. & Callaghan, T. V. Effects of increasing UV-B radiation on decomposition and soil organic matter dynamics. A synthesis and modeling study. *Biol. Fertil. Soils* **18**, 19–26 (1994).
- Pancotto, V. A. *et al.* Solar UV-B decreases decomposition in herbaceous plant litter in Tierra del Fuego, Argentina: potential role of an altered decomposer community. *Glob. Change Biol.* **9**, 1465–1474 (2003).
- Sala, O. E., Golluscio, R. A., Lauenroth, W. K. & Soriano, A. Resource partitioning between shrubs and grasses in the Patagonian steppe. *Oecologia* **81**, 501–505 (1989).
- Yahdjian, L., Sala, O. E. & Austin, A. T. Differential controls of water input on litter decomposition and nitrogen dynamics in the Patagonian steppe. *Ecosystems* **9**, 128–141 (2006).
- Smith, S. D. *et al.* Elevated CO<sub>2</sub> increases productivity and invasive species success in an arid ecosystem. *Nature* **408**, 79–82 (2000).
- Jackson, R. B., Banner, J. L., Jobbágy, E. G., Pockman, W. T. & Wall, D. H. Ecosystem carbon loss with woody plant invasion of grasslands. *Nature* **418**, 623–626 (2002).
- Thomas, D. S. G., Knight, M. & Wiggs, G. F. S. Remobilization of southern African desert dune systems by twenty-first century global warming. *Nature* **435**, 1218–1220 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge the late A. Soriano for establishment of a research program at the INTA study site more than 50 years ago; C. Mazza, L. Raiger, P. Flombaum, N. Sala, J. Vrsalovic, P. Araujo, L. Gherardi, M. Gonzalez-Polo, V. Marchesini, A. Fernández-Souto, P. Rojas, M. Tagliacchi and L. Yahdjian for field and laboratory assistance; and O. Sala, P. Vitousek, K. O'Shea, G. Piñeiro and C. Ballaré for comments on the manuscript. We acknowledge financial support from the Fundación Antorchas and the Fundación YPF of Argentina, the Inter-American Institute for Global Change Research, the US National Science Foundation, the Agencia Nacional de Promoción de Ciencia y Tecnología (ANPCyT) and the Universidad de Buenos Aires (UBACyT) of Argentina.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.T.A. ([austin@ifeva.edu.ar](mailto:austin@ifeva.edu.ar)).

# Record of mid-Archaeon subduction from metamorphism in the Barberton terrain, South Africa

Jean-François Moyen<sup>1</sup>, Gary Stevens<sup>1</sup> & Alexander Kisters<sup>1</sup>

Although plate tectonics is the central geological process of the modern Earth, its form and existence during the Archaeon era (4.0–2.5 Gyr ago) are disputed<sup>1,2</sup>. The existence of subduction during this time is particularly controversial because characteristic subduction-related mineral assemblages, typically documenting apparent geothermal gradients of 15 °C km<sup>-1</sup> or less<sup>3</sup>, have not yet been recorded from *in situ* Archaeon rocks (the lowest recorded apparent geothermal gradients<sup>4</sup> are greater than 25 °C km<sup>-1</sup>). Despite this absence from the rock record, low Archaeon geothermal gradients are suggested by eclogitic nodules in kimberlites<sup>5,6</sup> and circumstantial evidence for subduction processes, including possible accretion-related structures<sup>2</sup>, has been reported in Archaeon terrains. The lack of spatially and temporally well-constrained high-pressure, low-temperature metamorphism continues, however, to cast doubt on the relevance of subduction-driven tectonics during the first 1.5 Gyr of the Earth's history<sup>7</sup>. Here we report garnet–albite-bearing mineral assemblages that record pressures of 1.2–1.5 GPa at temperatures of 600–650 °C from supracrustal amphibolites from the mid-Archaeon Barberton granitoid–greenstone terrain. These conditions point to apparent geothermal gradients of 12–15 °C—similar to those found in recent subduction zones—that coincided with the main phase of terrane accretion in the structurally overlying Barberton greenstone belt<sup>8</sup>. These high-pressure, low-temperature conditions represent metamorphic evidence for cold and strong lithosphere, as well as subduction-driven tectonic processes, during the evolution of the early Earth.

Recent studies have highlighted the composite nature of the early-to mid-Archaeon Barberton granitoid–greenstone terrain and have demonstrated that the deep crustal levels (30–40 km) are exposed in structurally bounded domains in the granitoid–gneiss terrain to the south of the shallow-crustal greenstone belt<sup>9,10</sup>. Sedimentological, structural and geochronological differences indicate that the belt is made up of a northern and a southern terrane that are separated by the central, NE–SW trending Saddleback–Inyoka fault. The amalgamation of these two proposed island-arc terranes occurred during the main, D2 phase of collisional tectonics at ~3.23 Gyr ago, probably in an arc-trench setting<sup>8</sup>. The surrounding, granitoid–greenstone terrain is made up of (1) an amphibolite-facies greenstone component, (2) mainly gneissic trondhjemitic plutonic rocks ~3.55–3.45 Gyr old, (3) syntectonic (D2) trondhjemitic and tonalites ~3.23–3.22 Gyr old<sup>11</sup>. The older trondhjemitic and the associated greenstone remnants form an extensive, relatively high-grade domain (the Stolzburg terrane, Fig. 1), that was metamorphosed to pressures of up to 8–11 kbar (refs 9, 10). Significantly, peak metamorphic conditions in the high-grade gneiss terrain were attained during the D2 phase of tectonism, coeval with the accretion of the two island-arc terranes in the shallow-crustal greenstone belt. In other words, the supracrustal greenstone belt and the deep-crustal gneiss terrain

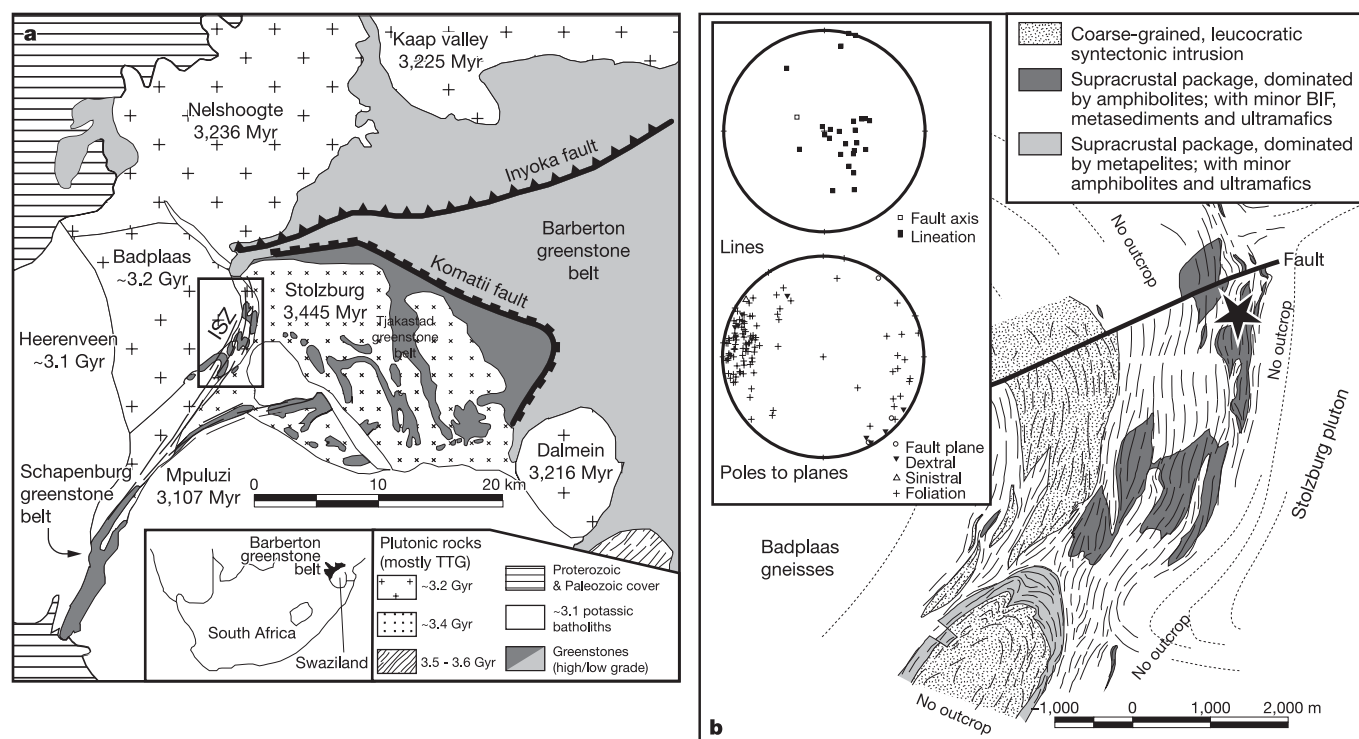
expose sections through different crustal levels of the ~3.23-Gyr collisional orogen.

This study presents the results of a metamorphic analysis of rare mineral assemblages found in supracrustal remnants from within a prominent shear zone within this gneiss terrain, the Inyoni shear zone, which is probably the mid- to lower crustal expression of the suture that accommodated the mid-Archaeon terrane accretion. The Inyoni shear zone is an up to 3 km wide, north-trending subvertical belt of banded, often migmatitic gneisses (Fig. 1). It extends southwards, to the kilometre-scale amphibolite-facies Schapenburg Schist belt. In the west, it is intruded by syntectonic bodies of coarse-grained, leucocratic trondhjemitic to granodiorites. Some of these D2 bodies yielded ages of  $3.229 \pm 0.005$  Gyr (ref. 12) and  $3.231 \pm 0.005$  Gyr (ref. 13), constraining the timing of the deformation. Towards the east, the zone is bounded by relatively homogeneous and lower-strain gneisses of the high-grade Stolzburg terrane. Although heterogeneous strain and high degrees of fabric transposition make it difficult to establish the overall kinematics and strain within the shear zone, the scarcity of non-coaxial fabrics and the fabric geometry point to a dominantly bulk flattening strain associated with a component of vertical extrusion of the rocks. The gneisses of the Inyoni shear zone contain metre- to kilometre-scale, variably deformed and metamorphosed metavolcanic and subordinate metasedimentary remnants. Lithological differences between mappable packages, both in the Inyoni shear zone proper, and in the Schapenburg Schist belt, together with contrasting pressure–temperature (*P–T*) conditions (8–11 kbar and 650–700 °C (ref. 14) in the North; ~5 kbar and 630 °C in Schapenburg) suggest that this composite gneiss belt is a tectonic melange, juxtaposing rocks from diverse crustal depths intruded by largely synkinematic granitoids. A northern metavolcanic package (Fig. 1) consists predominantly of layered, epidote- and hornblende-dominated amphibolites. Garnet occurs within specific, relatively iron-rich horizons and the metamorphic history of this zone can best be understood by focusing on these pressure-sensitive garnet-bearing assemblages.

Prograde metamorphic evolution is recorded in low-strain domains, such as the cores of rootless isoclinal folds, where garnet grew simultaneously with albitic plagioclase, as evidenced by euhedral garnets surrounded by plagioclase (Fig. 2a) or albitic inclusions within garnets, sometimes with negative garnet forms. Clinopyroxene and quartz are sometimes intergrown with garnet (Fig. 2b). This assemblage formed at the expense of a relatively sodic amphibole (Fe–edenite, up to 1.1 sodium atoms per formula unit), and epidote, partially reequilibrated relicts of which are found in crystallographic continuity within albitic moats around the garnets. Qualitatively, garnet–clinopyroxene–quartz assemblages are known to form at relatively high pressures<sup>15</sup>. In coexisting garnet–plagioclase pairs, Ca is preferentially partitioned into garnet over plagioclase as pressure increases<sup>15</sup>; thus, relatively calcic garnets coexist with sodic

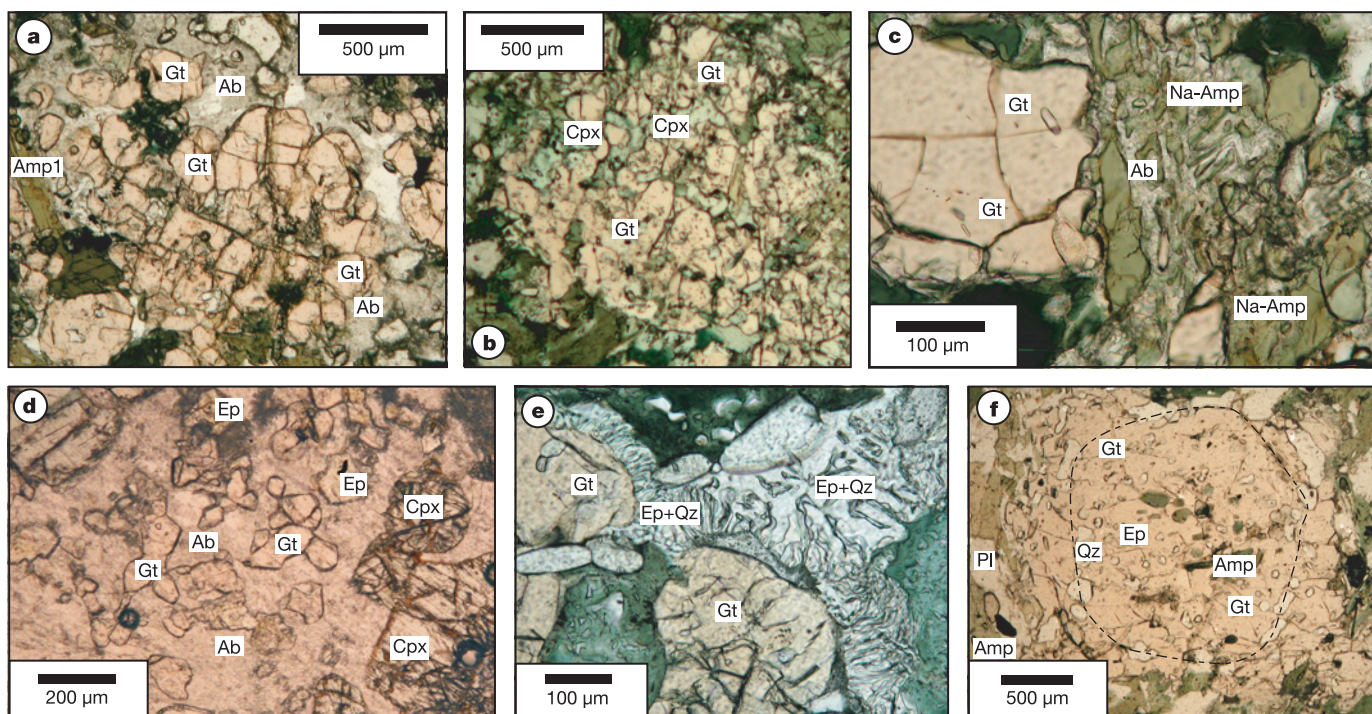
<sup>1</sup>Department of Geology, University of Stellenbosch, South Africa Private Bag X-1, Matieland 7602, South Africa.





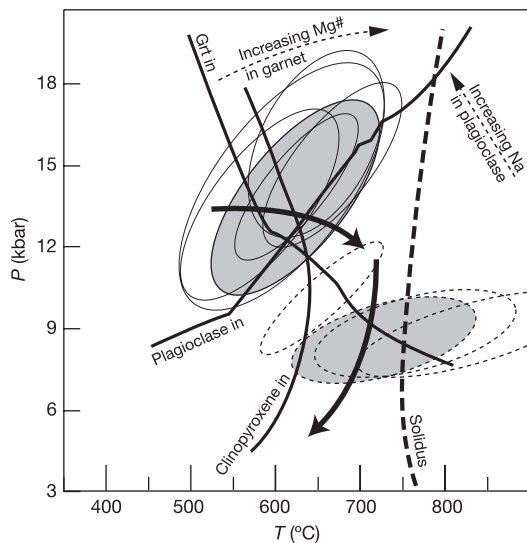
**Figure 1 | Location of the Inyoni shear zone and studied samples in the southern Barberton terrane<sup>14</sup>.** **a**, Regional geological map. Ages are from refs 11 and 21. ISZ, Inyoni shear zone. The box indicates the location of the detailed map in **b**. The darker-grey domain and the areas marked with smaller crosses correspond to the high-grade ~3.45-Gyr Stolzburg block. **b**, Detailed geological map of the northern Inyoni shear zone. The white

background represents the felsic, trondhjemitic and tonalitic (TTG) gneisses. Dotted lines display foliation trends. BIF, banded iron formation. Inset, stereograms (Schmidt, lower hemisphere) for the central domains of the mapped area. The star in **b** denotes the location of the studied samples, which were derived from different amphibolite bodies.



**Figure 2 | Metamorphic textures associated with garnet growth or breakdown (plane-polarized light).** **a**, Small euhedral garnets rimmed by albitic plagioclase, coalescing into large, poekilitic grains (INY131). **b**, Garnet-clinopyroxene intergrowth associated with garnet formation; the large garnet-clinopyroxene grain is rimmed by albite, not seen at this magnification (INY115). **c**, Breakdown of sodic amphibole within the albitic moats rimming garnets (INY131). **d**, Albitic plagioclase with euhedral grains of garnet and epidote, evidence of the breakdown of an earlier, more calcic

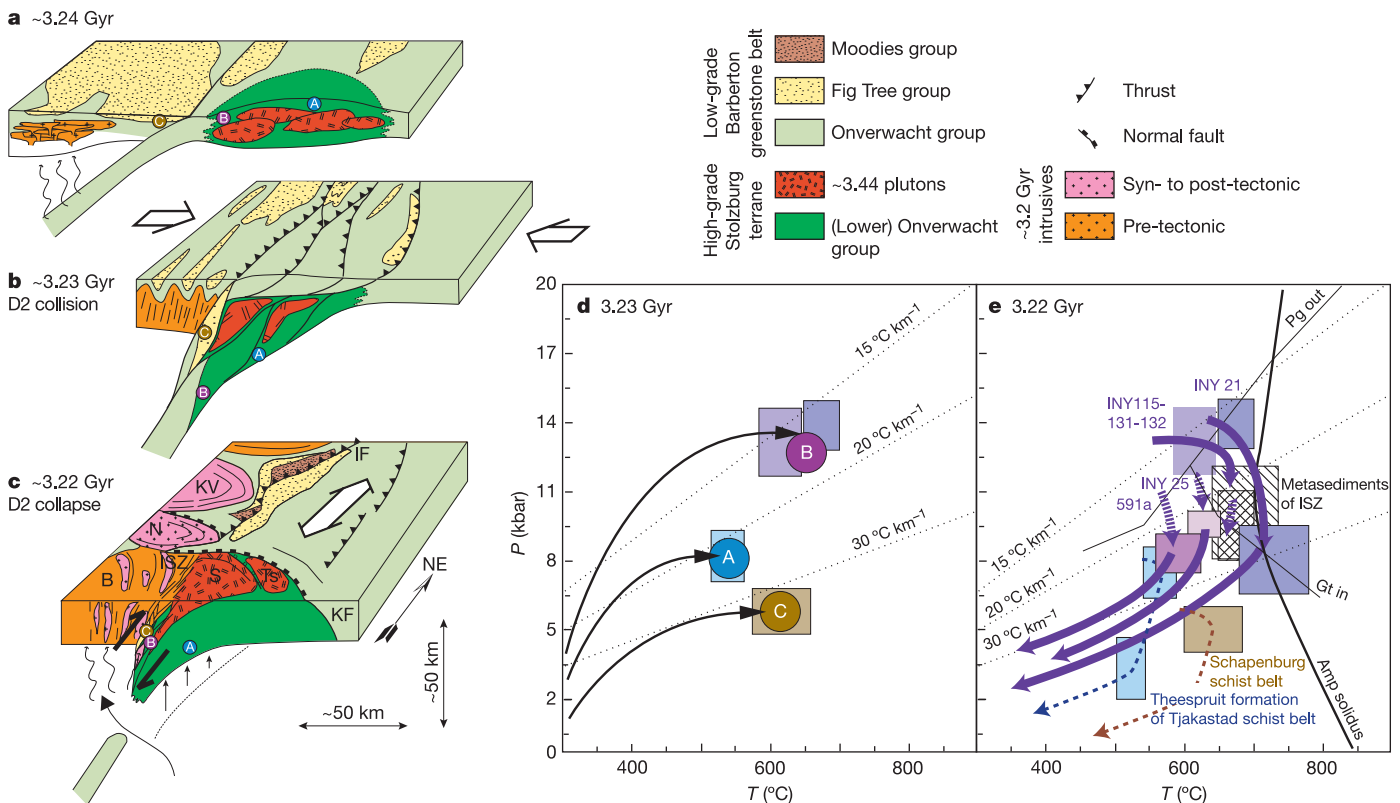
plagioclase. **e**, Garnet surrounded by epidote-quartz symplectites (INY25). **f**, Core-to-rim zoned garnet, with a low-temperature core with albite, epidote and amphibole inclusions rimmed by a higher-temperature rim in equilibrium with more calcic plagioclase and amphibole; a ring of quartz inclusions bounds the core (dashed line) (INY21). Amp, amphibole (Amp1, early Na-rich amphibole); Pl, plagioclase (Ab, albitic plagioclase); Gt, garnet; Cpx, clinopyroxene; Ep, epidote; Qz, quartz.



**Figure 3 | THERMOCALC  $P$ - $T$  estimates for garnet growth and breakdown sites in the studied samples.** Filled ellipses are the average for 11 (garnet growth, solid lines) and 9 (garnet breakdown, dashed lines) sites; empty ellipses are examples of individual calculations, each corresponding to one single reaction site using the compositions of minerals in textural equilibrium. Different mineral assemblages were used (for example, garnet-clinopyroxene-plagioclase-quartz, and garnet-amphibole-plagioclase-epidote-quartz, for the garnet growth sites); they all produce indistinguishable  $P$ - $T$  estimates within error. Imprecision associated with the plagioclase activity model, especially for low-An contents<sup>22</sup>, creates a relative error on An activity of 5–15% in garnet growth sites (low-An plagioclase) and 2–5% in garnet breakdown sites (intermediate plagioclase composition). An additional source of error is that inherent to the calibration of the garnet-clinopyroxene-plagioclase-quartz barometer<sup>22</sup>. These uncertainties are integrated within the THERMOCALC  $P$ - $T$  estimates<sup>23,24</sup>. Despite the significant error ellipses, the uncertainty on the apparent geotherm is much lower, owing to the positive slope of the reactions used in  $P$ - $T$  estimates.

plagioclase at the highest pressures of plagioclase-and-garnet coexistence (Fig. 3). The prograde garnet generation documented in this study is indeed calcic (35–40% grossular) and coexists with almost pure sodic endmember albitic plagioclase (An<sub>3–10</sub>). The garnet-in reaction is steeply orientated in  $P$ - $T$  space in the area where it intersects the high-pressure plagioclase phase boundary. The magnesium number Mg# (Mg/Mg + Fe) in garnet scales inversely with temperature away from this phase boundary. The low Mg# (10–15) of the prograde garnet in these rock compositions (Mg# ≈ 50) argues for

low-temperature garnet formation (Fig. 3). This approach is confirmed by estimates using THERMOCALC<sup>16</sup> thermobarometry (analytical techniques and representative mineral estimates are given in Supplementary Tables 1 and 2, respectively), consistently pointing to conditions of formation of 12–15 kbar and 600–650 °C (Fig. 3) for the garnet-bearing assemblage. Garnet from samples in the high-strain domains (samples INY 25, 591a) generally shows retrograde textures (Fig. 2e). Garnet breakdown conditions, recorded by epidote + Fe-tschermakite + quartz symplectite coronas around



**Figure 4 | Geodynamic sketch summarizing the inferred tectonic evolution of the southern Barberton terrane, and the associated metamorphic evolution.** a–c, Circles lettered A, B and C correspond to the Theespruit (Ts) formation of the Tjakastad schist belt<sup>10</sup>, the ISZ samples from this study, and the Schapenburg greenstone belt<sup>13</sup>, respectively. IF, Inyoka fault; KF,

Komatij fault. Plutons are: Ts, Theespruit; KV, Kaap valley; N, Nelshoogte; B, Badplaas; S, Stolzberg. **d, e**, The  $P$ - $T$  of points A, B and C during the assembly and collapse phases of the orogen are shown. The two hatched blocks marked 'metasediments of ISZ' in **e** are from ref. 9.



the garnets, correspond to (THERMOCALC) temperatures of 580–650 °C at 8–10 kbar. This set of metamorphic conditions is consistent with the position of the (negatively sloped) garnet phase boundary in this part of the  $P$ – $T$  space (Fig. 3); the estimated metamorphic conditions from these decompression structures corresponds well with peak metamorphic estimates from the nearby clastic sedimentary intercalations within the metavolcanic sequence<sup>9</sup>. The peak pressure  $P$ – $T$  estimates are at present the highest crustal pressures reported for Archaean rocks, and correspond to by far the lowest known apparent geothermal gradients ( $\sim 12$  °C km<sup>−1</sup>) in the Archaean rock record. In the modern Earth, the only process capable of producing crustal rock evolution through this  $P$ – $T$  domain occurs within subduction zones.

The Inyoni shear zone is the structurally and lithologically composite western boundary of the structurally coherent, high-pressure, low-temperature Stolzberg granitoid-gneiss terrane. The presence of rocks with a high-pressure history consistent with a subduction origin in this zone suggests that this may conceivably represent the suture along which the high-grade continental Stolzberg terrane was rapidly buried to depths of at least 35–40 km (Fig. 4) (refs 9, 17). We suggest that the *mélange*-like character of the shear zone is the result of the structural imbrication of deeply buried slivers during the buoyancy-assisted return flow between or close to the downgoing slab and the overriding plate<sup>18</sup>. The abundance of synkinematic trondhjemites in the shear zone is likely to be the result of decompression melting of amphibolites during the retrograde exhumation path. The presence of these melts is possibly important to understanding the documented metamorphic signature. High-strain fabrics confined to synkinematic trondhjemites point to strain localization into the melts, which, in turn, is likely to assist the buoyancy- or extrusion-related exhumation of the rocks. The advective heat transfer associated with the intrusion of these synkinematic magmas also contributes to the syn- to late-collisional heat budget of the collisional belt that acted to partially destroy the evidence for the earlier high-pressure, low-temperature metamorphism. In many respects, this is similar to high-pressure amphibolites from more recent subduction–collision belts, which often occur as partially retrogressed boudins within migmatites<sup>19,20</sup>. Our findings of high-pressure, low-temperature metamorphic mineral assemblages from an active margin setting strongly suggest that lithospheric subduction was functioning as early as 3.2 Gyr ago before present.

Received 17 March; accepted 12 June 2006.

- Hamilton, W. B. Archaean magmatism and deformation were not products of plate tectonics. *Precamb. Res.* **91**, 143–179 (1998).
- de Wit, M. J. On Archaean granites, greenstones, cratons and tectonics: does the evidence demand a verdict? *Precamb. Res.* **91**, 181–226 (1998).
- Ernst, W. G. Tectonic history of subduction zones inferred from retrograde blueschist  $P$ – $T$  paths. *Geology* **16**, 1081–1084 (1988).
- Riciputi, L. R., Valley, J. W. & McGregor, V. R. Conditions of Archaean granulite-facies metamorphism in the Gothåb-Fiskenaeset region, southern West Greenland. *J. Metamorph. Geol.* **8**, 171–190 (1990).
- Rollinson, H. Eclogite xenoliths in West African kimberlites as residues from Archaean granitoid crust formation. *Nature* **389**, 173–176 (1997).
- Ireland, T. R., Rudnick, R. L. & Spetius, Z. Trace elements in diamond inclusions from eclogites reveal link to Archaean granites. *Earth Planet. Sci. Lett.* **121**, 199–213 (1994).
- Bjørnerud, M. G. & Austrheim, H. Inhibited eclogite formation: the key to the rapid growth of strong and buoyant Archaean continental crust. *Geology* **32**, 765–768 (2004).
- De Wit, M. J. et al. Formation of an Archaean continent. *Nature* **357**, 553–562 (1992).
- Dziggel, A., Stevens, G., Poujol, M., Anhaeusser, C. R. & Armstrong, R. A. Metamorphism of the granite–greenstone terrane south of the Barberton greenstone belt, South Africa: an insight into the tectono-thermal evolution of the 'lower' portions of the Onverwacht group. *Precamb. Res.* **114**, 221–247 (2002).
- Diener, J., Stevens, G., Kisters, A. F. M. & Poujol, M. Metamorphism and exhumation of the basal parts of the Barberton greenstone belt, South Africa: constraining the rates of mid-Archaean tectonism. *Precamb. Res.* **143**, 87–112 (2005).
- Kamo, S. L. & Davis, D. W. Reassessment of Archaean crustal development in the Barberton mountain land, South-Africa, based on U–Pb dating. *Tectonics* **13**, 167–192 (1994).
- Dziggel, A., Armstrong, R. A., Stevens, G. & Nasdala, L. Growth of zircon and titanite during metamorphism in the granitoid-gneiss terrain south of the Barberton greenstone belt, South Africa. *Mineral. Mag.* **69**, 1021–1038 (2006).
- Stevens, G., Droop, G. T. R., Armstrong, R. A. & Anhaeusser, C. R. Amphibolite-facies metamorphism in the Schapenburg schist belt: a record of the mid-crustal response to  $\sim 3.23$  Ga terrane accretion in the Barberton greenstone belt. *S. Afr. J. Geol.* **105**, 271–284 (2002).
- Kisters, A. F. M., Stevens, G., Dziggel, A. & Armstrong, R. A. Extensional detachment faulting and core-complex formation in the southern Barberton granite–greenstone terrain, South Africa: evidence for a 3.2 Ga orogenic collapse. *Precamb. Res.* **127**, 355–378 (2003).
- Kohn, M. J. & Spear, F. S. Empirical calibration of geobarometers for the assemblage garnet + hornblende + plagioclase + quartz. *Am. Mineral.* **74**, 77–84 (1989).
- Holland, T. J. B. & Powell, R. An internally consistent thermodynamic dataset for phases of petrological interest. *J. Metamorph. Geol.* **16**, 309–343 (1998).
- Diener, J. F. A., Stevens, G. & Kisters, A. F. M. High-pressure low-temperature metamorphism in the southern Barberton granitoid-greenstone terrain, South Africa: a record of overthickening and collapse of mid-Archaean continental crust. In *Archaean Geodynamics And Environments* (eds Benn, K., Mareschal, J.-C. & Condie, K.) 239–354 (AGU Geophysical Monograph Series Vol. 164, AGU, Washington, 2005).
- Chemenda, A. I., Mattauer, M. & Bokun, A. N. Continental subduction and a mechanism for exhumation of high-pressure metamorphic rocks: new modelling and field data from Oman. *Earth Planet. Sci. Lett.* **143**, 173–182 (1996).
- Nicollet, C. & Leyreloup, A. Pétrologie des niveaux trondhémiques de haute pression associés aux élogites et amphibolites des complexes leptyno amphiboliques du Massif Central français. *Can. J. Earth Sci.* **15**, 695–707 (1978).
- Bodinier, J. L., Burg, J.-P., Leyreloup, A. & Vidal, H. Reliques d'un bassin d'arrière arc subducté puis obducté dans la région de Marvejols (Massif Central). *Bull. Soc. Geol. Fr.* **8**, 20–34 (1988).
- de Ronde, C. E. J. & Kamo, S. L. An Archaean arc-arc collisional event: a short-lived (ca 3 Myr) episode, Weltvreden area, Barberton greenstone belt, South Africa. *J. Afr. Earth Sci.* **30**, 219–248 (2000).
- Spear, F. S. *Metamorphic Phase Equilibria and Pressure-Temperature-Time Paths* 535 (Mineralogical Society of America, Washington, 1993).
- Powell, R., Holland, T. J. B. & Worley, B. Calculating phase diagrams involving solid solutions via non-linear equations, with examples using THERMOCALC. *J. Metamorph. Geol.* **16**, 577–588 (1998).
- Holland, T. J. B. & Blundy, J. Non-ideal interactions in calcic amphiboles and their bearing on amphibole-plagioclase thermometry. *Contrib. Mineral. Petrol.* **116**, 433–447 (1994).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** J.-F.M.'s post-doctoral stay at Stellenbosch university is funded by the South African National Research Foundation (NRF) and by a bursary from the Department of Geology, Stellenbosch University. Running costs were provided by the NRF. We thank G. Droop and J. Bédard for reviews of earlier versions of this manuscript.

**Author Contributions** J.-F.M. and G.S. contributed equally to the metamorphic and petrologic analysis. All authors contributed to the interpretation of these results within the Barberton geodynamic framework.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.-F.M. (moyen@sun.ac.za) or G.S. (gs@sun.ac.za).



# The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches

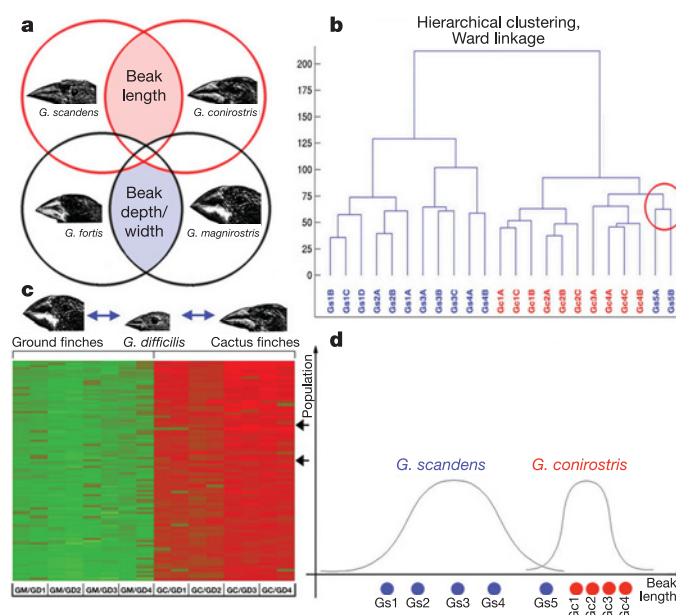
Arhat Abzhanov<sup>1†</sup>, Winston P. Kuo<sup>1,2,3†</sup>, Christine Hartmann<sup>4</sup>, B. Rosemary Grant<sup>5</sup>, Peter R. Grant<sup>5</sup> & Clifford J. Tabin<sup>1</sup>

A classic textbook example of adaptive radiation under natural selection is the evolution of 14 closely related species of Darwin's finches (Fringillidae, Passeriformes), whose primary diversity lies in the size and shape of their beaks<sup>1–6</sup>. Thus, ground finches have deep and wide beaks, cactus finches have long and pointed beaks (low depth and narrower width), and warbler finches have slender and pointed beaks, reflecting differences in their respective diets<sup>6</sup>. Previous work has shown that even small differences in any of the three major dimensions (depth, width and length) of the beak have major consequences for the overall fitness of the birds<sup>3–7</sup>. Recently we used a candidate gene approach to explain one pathway involved in Darwin's finch beak morphogenesis<sup>8</sup>. However, this type of analysis is limited to molecules with a known association with craniofacial and/or skeletogenic development. Here we use a less constrained, complementary DNA microarray analysis of the transcripts expressed in the beak primordia to find previously unknown genes and pathways whose expression correlates with specific beak morphologies. We show that calmodulin (CaM), a molecule involved in mediating  $\text{Ca}^{2+}$  signalling, is expressed at higher levels in the long and pointed beaks of cactus finches than in more robust beak types of other species. We validated this observation with *in situ* hybridizations. When this upregulation of the CaM-dependent pathway is artificially replicated in the chick frontonasal prominence, it causes an elongation of the upper beak, recapitulating the beak morphology of the cactus finches. Our results indicate that local upregulation of the CaM-dependent pathway is likely to have been a component of the evolution of Darwin's finch species with elongated beak morphology and provide a mechanistic explanation for the independence of beak evolution along different axes. More generally, our results implicate the CaM-dependent pathway in the developmental regulation of craniofacial skeletal structures.

To understand the genetic basis of the species-specific beak morphologies, we previously performed a comparative candidate gene analysis with developmental genes known to be associated with craniofacial development. We found that a broader and earlier domain of bone morphogenetic protein 4 (BMP4) expression in the distal neural-crest-derived mesenchyme correlated with the very deep and wide beak morphology of the ground finches<sup>8</sup>. This expression difference was shown to be functionally significant by misexpression analysis in chick embryos<sup>8</sup>.

However, the candidate gene approach did not yield any candidates for pathways that could be involved in evolution of the longer beak morphology characteristic of the cactus finch species. To identify pathways involved in the evolution of long beaks, cDNA

microarrays were used for a direct comparison of the gene expression profiles of several thousand transcripts in stage 26 frontonasal processes (which give rise to the upper beak) of five species of genus *Geospiza*: the sharp-beaked finch (*Geospiza difficilis*), the medium and large ground finches (*G. fortis* and *G. magnirostris*), and the cactus and large cactus finches (*G. scandens* and *G. conirostris*) (Fig. 1a; Methods). We first used hierarchical clustering to inspect whether the overall expression profiles clustered according to species (Methods). The resultant tree illustrates that most of the individual



**Figure 1 | Microarray analysis in different finch species.** **a**, Clustering strategy to isolate transcripts whose expression correlated with beak morphology. **b**, The Ward linkage tree showed that most of the individual samples clustered by species. Each individual was sampled two to four times. The y-axis is euclidian distance between branches. **c**, The final clusters of transcripts, which were upregulated in the comparison between cactus finches and the sharp-beaked finch and were downregulated or remained unchanged in the ground finches compared with the sharp-beaked finch. **d**, Individual expression profiles clustered by species except for the occasional individual profile, probably reflecting a certain overlap in morphology or development between the species. Each spot represents the length of an individual beak.

<sup>1</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>2</sup>Decision Systems Group, Brigham and Women's Hospital, Harvard Medical School, 310 Thorn Building, 75 Francis Street, Boston, Massachusetts 02115, USA. <sup>3</sup>Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, 188 Longwood Avenue, Boston, Massachusetts 02115, USA. <sup>4</sup>Institute of Molecular Pathology, Dr Bohr-Gasse 7, A-1030, Vienna, Austria. <sup>5</sup>Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA. <sup>†</sup>Present address: Department of Developmental Biology, Harvard School of Dental Medicine, Boston, Massachusetts 02115, USA.

expression profiles clustered by species. For example, most individuals of either *G. scandens* (4 of 5) or *G. conirostris* (4 of 4) clustered together with individuals of the same species (Fig. 1b). The expression profile of one particular individual of *G. scandens* (G.s.5) clustered more closely with those of *G. conirostris*, probably reflecting a certain degree of morphological (and thus developmental) overlap between these two species (Fig. 1d). *G. scandens* and *G. conirostris* collected for this study live on different islands (Santa-Cruz and Genovesa, respectively), thus excluding the possibility of species misidentification. We obtained very similar results when we compared *G. fortis* and *G. magnirostris* individuals (largely distinct expression profiles but some individuals clustering with the other species; not shown). It therefore seems that the analysis is sensitive to phylogenetic differences between different species of Darwin's finches.

We then clustered the measurements of signal ratios and intensities for different transcripts to identify genes that were upregulated or downregulated in all individuals of a particular morphology compared with the basal *G. difficilis* reference (Fig. 1, and Supplementary Fig. S1; Methods). All of 100 candidates from the resulting final cluster of cactus finch morphology-specific genes were sequenced to reveal their identity. These genes were screened to produce candidates that were expressed at moderate or high levels of signal intensity on the microarray and were expressed at least fivefold higher in the cactus finches. We found two microarray spots carrying probes with identical sequences for CaM (Supplementary Fig. S2) that were both at a much higher level of average signal in the cactus finch beaks than in the reference sharp-beaked finch (*G. difficilis*) (Supplementary Information). Because these clones represented the most differentially expressed gene that was not an enzyme, a house-keeping gene or a ribosomal gene and the only representative of a signalling transduction pathway, we focused on CaM as our primary candidate associated with the elongated cactus finch beak morphology. CaM is a  $\text{Ca}^{2+}$ -binding protein that can bind to and regulate many different protein targets and is a key component of a  $\text{Ca}^{2+}$ -dependent signal transduction pathway<sup>9</sup>. It has not been previously characterized in either craniofacial or skeletal development.

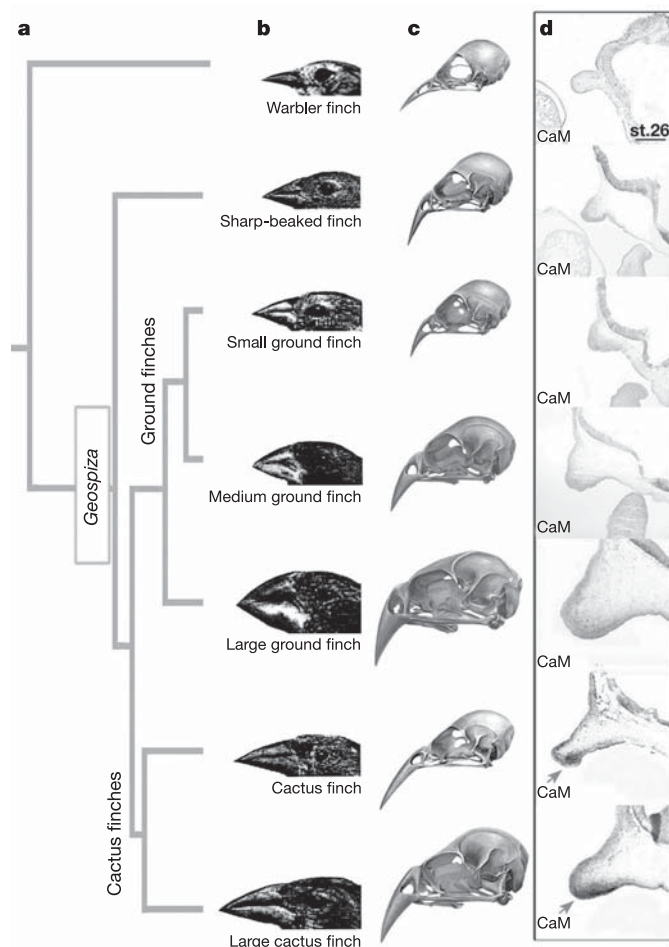
Our microarray data indicate a possible correlation between elevated levels of CaM and the more elongated beak morphology of the cactus finches. To validate this suggestion we performed a comparative *in situ* hybridization on embryos of Darwin's finches (Fig. 2). We found that CaM was indeed expressed at detectably higher levels in the distal-ventral mesenchyme of the frontonasal processes in the cactus finches ( $n = 3$ ) and large cactus finches ( $n = 3$ ) than in similar-sized processes of the ground finches (Fig. 2). Thus, higher expression of CaM is indeed associated with the elongated beak morphology of the cactus finches.

To address whether differential levels of CaM-dependent signalling might be important in the development of distinct beak morphologies in different species of Darwin's finches, we wished to elevate the level of CaM-dependent signalling in the developing chick beak prominence. To accomplish this we used a constitutively activated form of a downstream effector of CaM, CaM kinase kinase (CaMKII; M. J. Taschner, S. Schnaiter and C.H., unpublished observations). In one of the major cellular responses to increased  $\text{Ca}^{2+}$  levels, CaM activates CaMKII. CaMKII, in turn, activates downstream kinases, enabling them to phosphorylate and hence activate various targets, such as the transcription factors cAMP-response-element-binding protein (CREB), serum response factor and CREB-binding protein.

Both CaM and CamKII are expressed at low levels in the mesenchyme of the chicken upper beak process (data not shown). To stimulate higher CaM-dependent CAMKII signalling, we used an avian retroviral vector carrying the constitutively active form of CaMKII (RCAS:CA-CaMKII) to infect the distal mesenchyme of early stage-24 chick frontonasal processes. Although virus injection was targeted to the distal upper beak prominence with relative ease,

the resultant infection was somewhat variable. In more than half of the cases, the infection was limited to the distal and ventral mesenchyme surrounding the developing cartilage and did not spread to the skeletal element (Fig. 3d;  $n = 9$  of 16; not shown). Because these infected regions closely approximate the domain of elevated CaM expression in the cactus finches (Fig. 2), we restricted our analysis to these embryos. In chicks in which activated CaMKII was misexpressed specifically in the distal/ventral mesenchyme, we observed a significant increase in the length of the beaks (length in arbitrary units:  $0.78 \pm 0.05$  ( $\pm$ s.d.);  $P < 0.0056$ ,  $n = 9$ ; more than 10% beak elongation relative to control embryos (Fig. 3a–d)), whereas the beak width and depth were not affected (Fig. 3, and Supplementary Fig. S3).

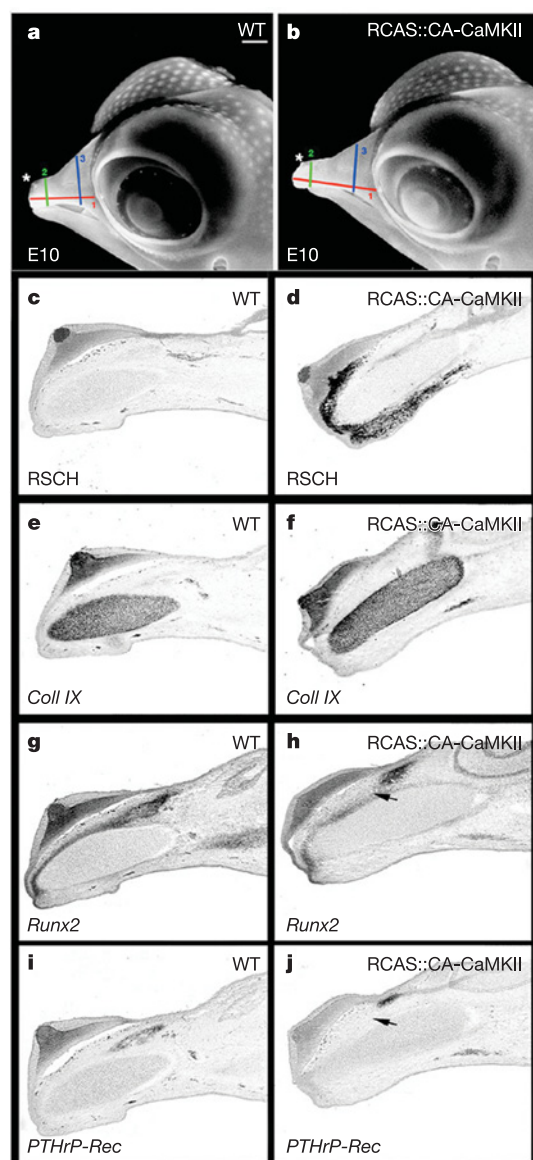
For a better understanding of the phenotype induced by activated CaMKII in beaks, we studied the expression of markers for skeletal cell types, proliferation and cell death (Fig. 3, and Supplementary Fig. S4). *In situ* hybridization analysis with probes directed against skeletogenic genes verified that the cartilage element was indeed longer in the distal/ventrally infected embryos (Fig. 3). We found that



**Figure 2 | Comparative analysis of CaM expression in finches.**

**a, b**, *Geospiza* group species displaying distinct beak morphologies form a monophyletic group. **c**, The differences in beak morphology are skeletal. **d**, CaM is expressed in a strong distal-ventral domain in the mesenchyme of the upper beak prominence of the large cactus finch, *G. conirostris*, somewhat lower levels in cactus finch, *G. scandens*, and at significantly lower levels in the large ground finch and medium ground finch, *G. magnirostris* and *G. fortis*, respectively. Very low levels of CaM were detected in the mesenchyme of *G. difficilis*, *G. fuliginosa* and the basal warbler finch *Certhidea olivacea*. CaM expression domains are indicated with short arrows in **d**. Scale bar, 1 mm in **b**. The molecular tree is from ref. 23; images of skulls are from ref. 6, with permission from the author.

cell death, as revealed by the TUNEL (TdT-mediated dUTP nick end labelling) assay, was essentially unchanged relative to the wild-type condition (not shown). Staining for anti-PCNA (proliferating cell nuclear antigen) antibody and labelling with bromodeoxyuridine indicate that proliferation might have been upregulated in the distal half of the pre-nasal cartilage element of infected beaks (not shown). The expression of cell-cycle regulators cyclin D1 and the activator protein 1 (AP1) family members c-Jun and c-Fos was upregulated in the distal parts of pre-nasal cartilage and its perichondrium (arrow in Supplementary Fig. S4j; not shown). As these skeletal tissues were not infected with the retrovirus, this effect was indirect. The same genes were downregulated in the mesenchyme, which forms the premaxillary dermal bone (arrowheads in Supplementary Fig. S4I).



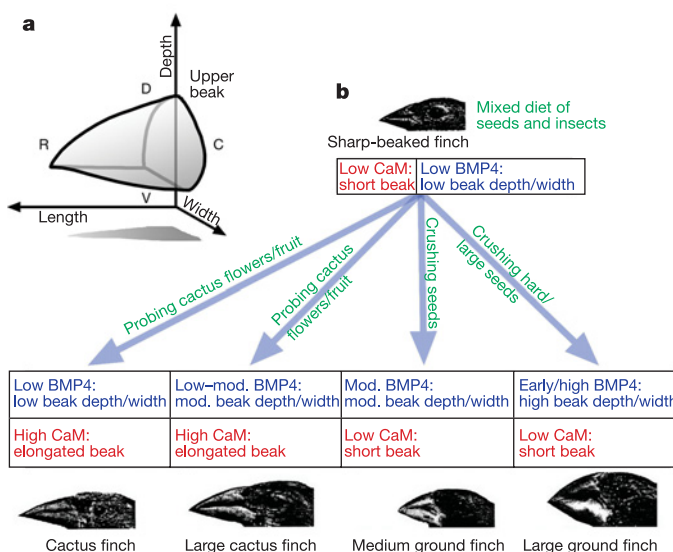
**Figure 3 | Functional analysis of CaM-dependent pathway in beak development.** **a, b**, Whole head views of embryonic day 10 (E10; HH stage 36) wild-type (**a**) and RCAS::CA-CaMKII-infected (**b**) chicken embryos. The length of the beak is shown with a red line; the depth of the beak at the base and the depth of the beak at the tip are shown with blue and green lines, respectively. **c–j**, We used RSCH (**c, d**), *Coll IX* (**e, f**), *Runx2* (**g, h**) and *PTHrP-Rec* (**i, j**) probes to reveal RCAS infection (RSCH), chondrocytes (*Coll IX*) and early osteoblasts (*Runx2*). **c, e, g, i**, Wild type; **d, f, h, j**, RCAS::CA-CaMKII infected. The star indicates an egg-tooth. Scale bar, 2 mm in **a**.

Consistent with our observations, parallel studies in the laboratory of one of us (C.H.) have also indicated that CaMKII signalling leads to the elongation of limb skeletal elements in a non-cell-autonomous manner (M. J. Taschner, S. Schnaiter and C.H., unpublished observations), in a similar manner to our observation in the beak.

There are few examples of the identification and characterization of developmental pathways responsible for evolutionary morphological change<sup>10–15</sup>. Previous work on Darwin's finches has shown that selection on adult beak variation results in changes in growth in the next generation, as a result of genetic correlations between adult and juvenile expression of the same morphological characters<sup>16,17</sup>. Here, using a microarray approach, we found evidence that a higher level of CaM-dependent signalling is both biologically relevant and functionally important for the morphogenesis of the longer beaks of cactus finches used in a specialized manner to probe cactus flowers and fruit for nectar.

The avian beak varies between species, and indeed between individuals of the same species, along at least three different axes: length, depth and width. In Darwin's finches it can be seen that there is both linkage and independence in the variation along these different axes. For example, in the two species of cactus finches there is a much weaker genetic correlation between adult beak length and depth or width than between depth and width themselves. In contrast, in *G. fortis* all three correlations are strong<sup>18,19</sup>. The data reported here provide at least a partial explanation for this difference between the finch species. Embryos of the cactus finch species, in which beak length shows independence from width and length, strongly express CaM (a factor specifically affecting beak length) during beak morphogenesis, whereas *G. fortis* has minimal CaM expression (Fig. 4).

In considering the independence of beak length from width and depth, it is particularly intriguing that, here and in our previous study, we have shown that two different factors (BMP4 and activated CaMKII), expressed in a similar domain in the beak prominence, result in changes in growth along different dimensions of the developing beak (width/depth and length, respectively) without negatively affecting



**Figure 4 | BMP- and CaM-dependent signalling regulates growth along different axes, facilitating the evolution of distinct beak morphologies in Darwin's finches.** **a**, Developing avian beak is a three-dimensional structure that can change along any of the growth axes. **b**, A beak of the sharp-beaked finch reflects a basal morphology for *Geospiza*. The model for BMP4 and CaM involvement explains development of both elongated and deep/wide beaks of the more derived species. Abbreviations: C, caudal; D, dorsal; R, rostral; V, ventral.



the other axes, explaining the observed independence of these traits. Moreover, analysis of BMP4 and CaM expression in chick embryos infected with RCAS::CA-CaMKII and RCAS::BMP4, respectively, showed that these two molecules do not regulate each other during beak development (data not shown). These two molecules vary independently of each other in some species as demonstrated by their reciprocal expression levels in the two medium-sized species with thick and with long beak morphologies, whereas they are expressed together in the developing beak of the large cactus finch in good correlation with its robust yet elongated beak morphology (Figs 1, 3 and 4)<sup>6,8</sup>. Theoretically, such modular developmental regulation need not have been necessary to construct a beak. A common set of growth-promoting pathways could, in principle, have led to outgrowth of the frontonasal process along all three dimensions. Indeed, a single factor, BMP4, seems to stimulate growth of the beak along two dimensions, promoting deeper and wider beaks, explaining the linkage in their variation. Using a common signal for outgrowth in different axes still permits evolution of form, for example through the action of localized antagonists. However, the developmental programme using a distinct pathway for the long axis of the beak enhances the 'evolvability' along this dimension, by allowing independent variation along the different axes. Thus, the dissociable regulation of growth in different beak axes is an example of the organization of the biology of the organism facilitating the generation of variation on which natural selection can act<sup>20</sup>.

It will be of interest to sample other species of Darwin's finches with elongated beaks such as the woodpecker finch *Cactospiza pallida*, as well as other avian species with long beaks, to discover how generally the CaM-dependent pathway is involved in elongation of the beak skeleton. Upregulation of this pathway at stage 26 is not simply related to low beak depth and width of beaks, because CaM expression was very low in the pointed but slender beaks of the warbler finch embryos (Fig. 3). Upregulation of CaMKII activity in the distal mesenchyme of the upper beak prominence in chicken embryos produced a roughly 10% increase in beak length, primarily due to elongation of the pre-nasal cartilage rod, which is in concert with the differences in relative beak length between the cactus finches and the more basal sharp-beaked finch and other ground finches<sup>6</sup>. This might indicate that additional pathways and mechanisms have a function in generating the even longer beaks found in other avian lineages, such as some shorebirds, hummingbirds and Hawaiian honeycreepers, all highly adaptive.

## METHODS

**Collection and treatment of embryonic material from Darwin's finches.** Under an agreement with the Galápagos National Park, we received quotas for collecting embryos of *G. magnirostris*, *G. fortis*, *G. fuliginosa* and *G. scandens* from Santa Cruz Island, and *G. conirostris*, *G. difficilis* and *C. olivacea* from Genovesa Island. To avoid causing nest defection, only the third egg was collected shortly after it was laid; it was then incubated at 100 °F (37.8 °C). Embryonic material was fixed in 4% paraformaldehyde in PBS for 2 h at ambient temperature and stored in RNAlater reagent (Ambion) at about 5 °C for two to five weeks. Chick antisense riboprobes (*CaM*, *Coll IX*, *Runx2* and *PTHrP-Rec*) were prepared and used on Darwin's finch embryos as described previously<sup>8</sup>. We analysed 26 heads of Darwin's finches: large ground finch ( $n = 3$ ), medium ground finch ( $n = 5$ ), small ground finch ( $n = 4$ ), large cactus finch ( $n = 4$ ), cactus finch ( $n = 3$ ), sharp-beaked finch ( $n = 3$ ) and warbler finch ( $n = 4$ ).

**Chicken embryo manipulations and statistical analysis.** Fertilized eggs (Spafas) were incubated at 100 °F. Embryos were staged as described in ref. 21. The RCAS::CA-CaMKII construct will be described elsewhere (M. J. Taschner, S. Schnaiter and C.H., unpublished observations). To infect embryos for studies *in vivo* we injected the distal part of the frontonasal process of HH stage-24 chick embryos. The infected embryos were collected at embryonic day 10, fixed overnight in 4% paraformaldehyde in PBS, and frozen in OCT (optimal cutting temperature) medium for sectioning. Extent of viral infection was assayed by hybridization *in situ* with a virus-specific RSCH probe. The embryonic heads were photographed and measured in NIH Image 1.62. These arbitrary measurements were used for the analysis of variance function (ANOVA toolbox; *t*-test) in Microsoft Excel X to calculate standard deviations and *P* values for the data.

**Darwin finch microarray production and usage.** A DNA microarray (21,168 spots) was printed from a non-normalized poly(A)-primed cDNA library made from RNA isolated from multiple (12 individuals) frontonasal processes of stage-26 and stage-29 embryos of the medium ground finch, *G. fortis*. We used Cy5-labelled probes made from individual frontonasal processes of the four derived species of *Geospiza* for direct comparisons against a common Cy3-labelled reference sample made from pooled RNA of several (nine individuals) embryos of more basal *G. difficilis* (Fig. 1a, and Supplementary Fig. S1). In most cases we compared four unrelated individuals from each of the derived cactus finch and ground finch species (*G. scandens*, *G. conirostris*, *G. magnirostris* and *G. fortis*) against the pooled common reference (Figs 1a and 4, and Supplementary Fig. S1). RNA from each individual finch beak prominence was independently amplified and labelled in triplicate with a control dye swap. We used the two highest-quality sets of microarray data from each triplicate for clustering. Raw.gpr files were generated with GenePix 3.0 (Molecular Devices). Normalization and statistical analysis of the GPR data files were performed in MatLab (The Math Works). Data were normalized with the Lowess algorithm<sup>22</sup>. Only spots with a signal intensity exceeding the median background +2 s.d. were considered, which left 7,369 spots. The data was log<sub>2</sub> transformed.

**Microarray cluster analysis.** Clustering analysis and visualization were computed in MatLab. Agglomerative hierarchical clustering was performed by using the euclidean distance measure: the average linkage and Ward heuristics were used to connect the gene clusters. For *k*-means clustering, the *k*-means algorithm partitioned the genes into *k* discrete clusters on the basis of their expression. The number *k* (50) was preselected. The resultant tree illustrates that duplicates of the amplification/labeling experiments from the same individuals clustered together (Fig. 1b).

Measurements of signal ratios and intensities for different transcripts were clustered to identify genes that were upregulated or downregulated in all individuals of a particular species compared with the basal *G. difficilis* reference. Species-specific clusters were further cross-compared to reveal transcripts that were consistently upregulated in frontonasal processes of all individuals of the cactus-finch beak morphology and that remained unchanged or were downregulated in beak primordia of the ground finches. Conveniently, the cactus finch has an upper beak depth similar in size to that of the medium ground finch, whereas the large cactus finch's beak is more similar in size to that of the large ground finch while differing strongly in shape. This allowed us to separate transcripts exhibiting the size-specific regulation from those with shape-specific regulation.

**Data analysis: hierarchical clustering.** Gene expression patterns of nine experimental samples representing five *G. scandens* and four *G. conirostris* samples were analysed by hierarchical clustering with Ward linkage. The samples were divided into two groups on the basis of differences in gene expression. Figure 2 shows a cluster dendrogram of the two groups, *G. scandens* (blue) and *G. conirostris* (red). Within each branch the siblings were found to cluster tightly together on a sub-branch.

Received 15 February; accepted 25 April 2006.

1. Darwin, C. R. *Journal of Researches into the geology and Natural History of the various countries visited during the voyage of H.M.S. Beagle, under the command of Captain FitzRoy, R.N.* 2nd edn (John Murray, London, 1845).
2. Lack, D. *Darwin's Finches* (Cambridge Univ. Press, Cambridge, 1947).
3. Grant, P. R. *The Ecology and Evolution of Darwin's Finches* (Princeton Univ. Press, Princeton, New Jersey, 1999).
4. Futuyma, D. J. *Evolutionary Biology* 3rd edn (Sinauer Associates, Sunderland, Massachusetts, 1998).
5. Freeman, S. & Herron, J. C. *Evolutionary Analysis* 3rd edn (Prentice Hall, Englewood Cliffs, New Jersey, 2003).
6. Bowman, R. I. Morphological differentiation and adaptation in the Galapagos finches. *Univ. Calif. Publ. Zool.* **58**, 1–302 (1961).
7. Grant, P. R. & Grant, B. R. Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**, 707–711 (2002).
8. Abzhanov, A., Protas, M., Grant, R. B., Grant, P. R. & Tabin, C. J. Bmp4 and morphological variation of beaks in Darwin's finches. *Science* **305**, 1462–1465 (2004).
9. O'Day, D. H. CaMBOT: profiling and characterizing calmodulin-binding proteins. *Cell. Signal.* **15**, 347–354 (2003).
10. Burke, A. C., Nelson, C. E., Morgan, B. A. & Tabin, C. Hox genes and the evolution of vertebrate axial morphology. *Development* **121**, 333–346 (1996).
11. Averof, M. & Patel, N. H. Crustacean appendage evolution associated with changes in Hox gene expression. *Nature* **388**, 682–686 (1997).
12. Abzhanov, A. & Kaufman, T. C. Novel regulation of the homeotic gene *Scr* associated with a crustacean leg-to-maxilliped appendage transformation. *Development* **126**, 1121–1128 (1999).
13. Abzhanov, A. & Kaufman, T. C. Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk. *Development* **127**, 2239–2249 (2000).

14. Shapiro, M. D. *et al.* Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).
15. Kimmel, C. B. *et al.* Evolution and development of facial bone morphology in threespine sticklebacks. *Proc. Natl. Acad. Sci. USA* **102**, 5791–5796 (2005).
16. Grant, P. R. Inheritance of size and shape in a population of Darwin's finches. *Proc. R. Soc. Lond. B* **212**, 403–432 (1983).
17. Price, T. D. & Grant, P. R. The evolution of ontogeny in Darwin's Finches: a quantitative genetics approach. *Am. Nat.* **125**, 169–188 (1985).
18. Grant, B. R. & Grant, P. R. *Evolutionary Dynamics of a Natural Population. The Large Cactus Finch of the Galápagos* (Univ. Chicago Press, Chicago, Illinois, 1989).
19. Grant, P. R. & Grant, B. R. Phenotypic and genetic effects of hybridization in Darwin's finches. *Evolution Int. J. Org. Evolution* **48**, 297–316 (1994).
20. Kirschner, M. W. & Gerhart, J. C. *The Plausibility of Life: Resolving Darwin's Dilemma* (Yale Univ. Press, New Haven, Connecticut, 2005).
21. Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92 (1951).
22. Dudoit, S., Yang, H., Callow, M. J. & Speed, T. P. Statistical methods for identifying genes with differential expression in replicated cDNA experiments. *Statist. Sin.* **12**, 111–139 (2002).
23. Petren, K., Grant, B. R. & Grant, P. R. A phylogeny of Darwin's finches based on microsatellite DNA length variation. *Proc. R. Soc. Lond. B* **266**, 321–329 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank all field assistants and participants of the field collecting trips—M. Protas, J. Chavez, G. Castaneda, O. Perez, F. Brown, A. Aitkhozhina, M. Gavilanes, M. Paez, K. Petren, J. Podos and S. Kleindorfer—for their advice with species identification and other advice; Charles Darwin Research Station on Santa Cruz Island and The Galápagos National Park for permits and logistical support; and M. Kirschner for discussions that led to the inception of this project. A.A. was supported by the Cancer Research Fund of a Damon Runyon–Walter Winchell Foundation Fellowship. This project was funded by a program project grant from the NIH to C.J.T.

**Author Contributions** A.A. performed embryonic material collection, microarray probe preparation, microarray hybridizations and scanning, sectioning of material, *in situ* hybridizations and CaMKII functional analysis in chicken embryos. W.P.K. conducted all relevant bioinformatics analyses. C.H. constructed the RCAS virus carrying the constitutively active version of CaMKII. B.R.G. and P.R.G. provided logistics and secured permits for the fieldwork on the Galápagos Islands. C.J.T. conceived and supervised the project. A.A., B.R.G., P.R.G. and C.J.T. co-wrote the manuscript. All authors discussed the results and commented on the manuscript.

**Author Information** The sequence of Darwin's finch CaM has been deposited at GenBank under accession number DQ386479, and the microarray data have been filed with ArrayExpress under the accession number E-MEXP-702. Reprints and permission information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondents and requests for materials should be addressed to C.J.T. ([tabin@genetics.med.harvard.edu](mailto:tabin@genetics.med.harvard.edu)).

## LETTERS

# Collinear activation of *Hoxb* genes during gastrulation is linked to mesoderm cell ingression

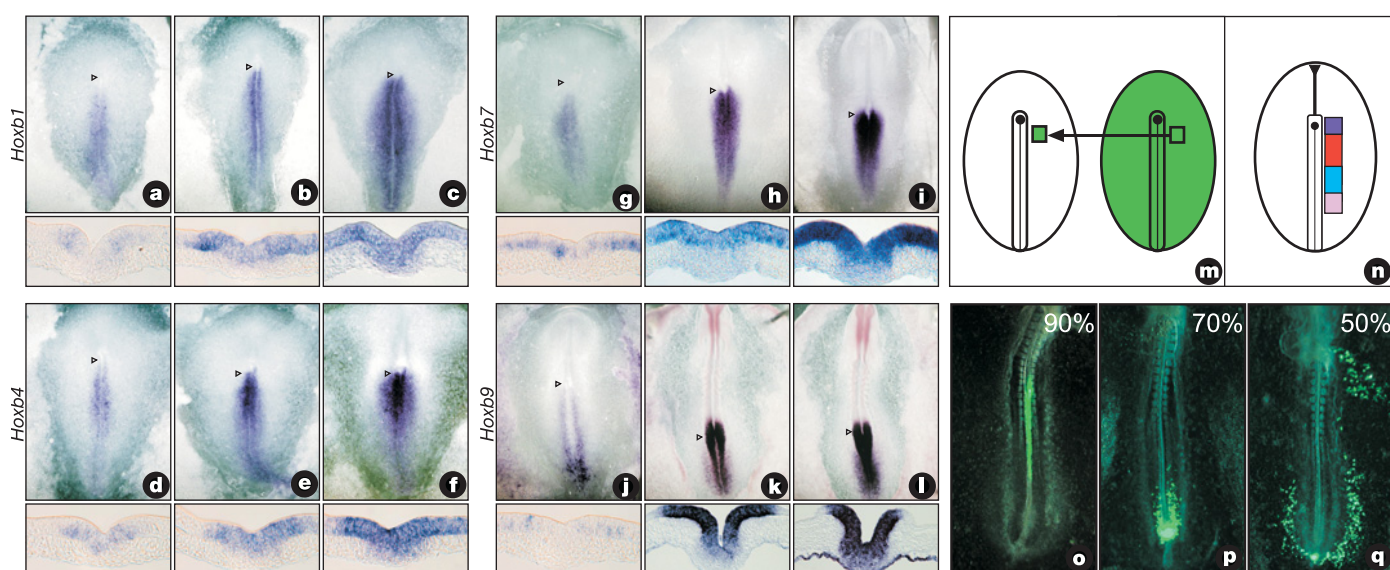
Tadahiro Iimura<sup>1</sup> & Olivier Pourquié<sup>1</sup>

The vertebral column exhibits segmentation and regionalization along the antero-posterior axis. During embryogenesis, the rhythmic production of the precursors of the vertebrae, the somites, imposes a segmented aspect to the spine, whereas the spine's regional differentiation is controlled by *Hox* genes<sup>1,2</sup>. Here we show that in the paraxial mesoderm, *Hoxb* genes are first activated in a temporal collinear fashion in precursors located in the epiblast lateral to the primitive streak. Our data suggest that collinear activation of *Hoxb* genes regulates the flux of cells from the epiblast to the streak and thus directly controls the establishment of the genes' characteristic nested expression domains in the somites. This suggests that establishment of the spatial co-linearity in the embryo is directly controlled by the *Hox* genes themselves.

The precursors of the vertebrae are segmented embryonic structures called somites. They derive from the presomitic mesoderm (PSM) located bilaterally on both sides of the neural tube in the posterior part of the embryo. At the anterior extremity of the PSM, pairs of somites form in a rhythmic fashion, thus establishing the metameric pattern of the embryonic axis. Depending on their location along the antero-posterior (AP) axis, somites will differentiate

according to specific morphogenetic programs, resulting in the establishment of distinct anatomical domains like the cervical, thoracic, lumbar and sacral regions. This regionalization of the vertebral column along the AP axis is largely controlled by a group of transcription factors called *Hox* genes. In mammals, these genes are organized in four clusters containing a total of 39 genes<sup>2</sup>. Within each cluster, the genes are arranged in a sequence that reflects their order of expression during development (temporal co-linearity<sup>3</sup>) and the position of the anterior boundary of their expression domains along the AP body axis (spatial co-linearity<sup>4,5</sup>). This nested distribution of *Hox* genes along the AP axis results in each vertebral precursor being endowed with a unique combinatorial expression of *Hox* genes that controls their AP identity<sup>6,7</sup>. In amniote embryos, activation of genes along the *Hox* clusters occurs sequentially in the primitive streak and the tail bud during axis elongation<sup>8</sup>.

We analysed the activation schedule of *Hoxb* genes in the paraxial mesoderm, in early chick embryos. *Hoxb1*, *Hoxb2*, *Hoxb3*, *Hoxb4*, *Hoxb7*, *Hoxb8* and *Hoxb9* were first detected in scattered cells in the epiblast adjacent to the posterior two-thirds of the primitive streak (Fig. 1a, d, g and j; top and bottom panels, Supplementary Fig. 1 and



**Figure 1 | *Hox* gene activation begins in the mesodermal territory of the epiblast.** **a–l**, Top panels, whole-mount *in situ* hybridization showing the activation of *Hoxb1* (**a**, stage 3+HH; **b**, 4HH; **c**, 5HH), *Hoxb4* (**d**, 4HH; **e**, 5HH; **f**, 6HH), *Hoxb7* (**g**, 5HH; **h**, 6HH; **i**, 7HH) and *Hoxb9* (**j**, 7HH; **k**, 5-somite; **l**, 6-somite). Bottom panels, sections of embryos shown in top panels at the 70% level of the streak. Dorsal views, anterior to the top. Arrowhead marks the 90% streak level. **m**, Homotopic and homochronic grafting of fragments of EGFP-electroporated epiblast lateral to the

primitive streak. **n**, Summary of the results of the epiblast fate mapping. Colours indicate the fate of the grafted tissue. Dark blue, neural; red, paraxial mesoderm; lighter blue, lateral plate mesoderm; pink, extraembryonic mesoderm. **o–q**, Embryos grafted as shown in **m** after 24 h of reincubation. A small fragment of labelled epiblast grafted adjacent to the 90% streak level mostly contributes to the neural tube (**o**); whereas, at the 70% level, cells are found in the paraxial mesoderm (**p**) and at the 50% level in the extraembryonic mesoderm and in the ectoderm (**q**). Ventral views, anterior to the top.

<sup>1</sup>Howard Hughes Medical Institute and Stowers Institute for Medical Research, 1000 E. 50th Street, Kansas City, Missouri 64110, USA.



data not shown). Double *in situ* hybridization with *Hoxb* genes and *Sox2*, which labels the neural plate, indicates that the genes are initially activated at a distance from the neural territory (Supplementary Fig. 1 and data not shown). The whole *Hoxb* cluster is activated in a collinear fashion in the epiblast in less than 12 hours between stages 4HH and 6HH (see Methods for stage nomenclature). Expression of the *Hoxb* genes subsequently appears to spread anteriorly and posteriorly to the entire lateral epiblast (Fig. 1b, e, h and k; top and bottom), reaching the level of the anterior primitive streak (arrowhead in Fig. 1b, e, h, k; top) and the posterior neural plate (Fig. 1b, e, h, k; top). Expression was then seen in cells ingressing into the primitive streak (Fig. 1c, f, i, l; top and bottom) and was maintained in the mesodermal derivatives produced by these expressing cells (not shown). Similar kinetics of *Hox* activation have been observed in the frog and mouse embryos, suggesting that this expression sequence is conserved across vertebrates<sup>9,10</sup>.

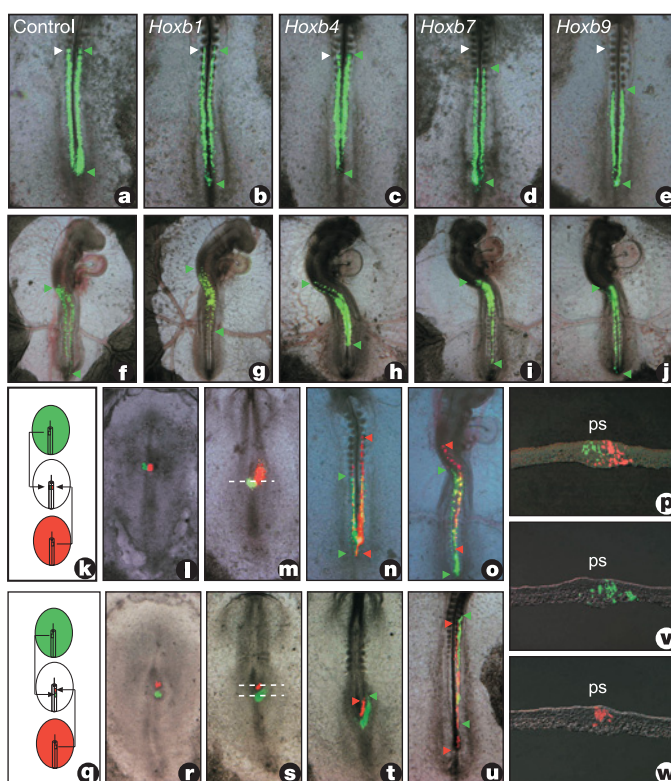
Because the fate of the epiblast adjacent to the primitive streak after the beginning of node regression is poorly characterized<sup>11–13</sup>, we performed a series of homotopic grafts of enhanced green fluorescent protein (EGFP)-labelled fragments of different levels of the epiblast adjacent to the primitive streak between stages 4HH and 7HH (Fig. 1m). The anteriormost epiblast, which lies adjacent to the node and the anterior primitive streak, essentially gave rise to neural and ectodermal derivatives (Fig. 1n, o,  $n = 14$ ). We found that the epiblast region lateral to the primitive streak at the 90% to 60% levels essentially contained paraxial mesoderm precursors (Fig. 1n, p,  $n = 25$ ). Posterior to the 60% level of the streak, the epiblast derivatives were essentially found in the lateral plate or extra-embryonic mesoderm (Fig. 1n, q,  $n = 11$ ). These observations indicate that ingression of somite precursor cells from the lateral epiblast continues long after the beginning of primitive streak regression. In the paraxial mesoderm, therefore, activation of the *Hoxb* cluster is first initiated in the somite precursors in the epiblast in a collinear fashion before their ingression into the primitive streak.

The activation kinetics of *Hox* genes in paraxial mesoderm precursors strikingly parallels the sequence of ingression of the epiblast into the primitive streak. To test if *Hox* genes control the ingression of epiblast cells into the streak, we grafted small fragments of the anterior streak overexpressing either *Hoxb1*, *Hoxb4*, *Hoxb7* or *Hoxb9* with a ZsGreen reporter into stage-matched embryos. We then scored the position of labelled descendants of grafted cells (Supplementary Table 1). In stage 5HH embryos grafted with an 80% primitive streak fragment overexpressing *Hoxb7* or *Hoxb9* constructs, labelled cells were found to extend more posteriorly than in embryos grafted with fragments overexpressing the *Hoxb1*, *Hoxb4* or a control construct (compare Fig. 2a–c with d, e). We then monitored the posterior distribution of labelled cells in grafted embryos incubated for a further 24-hour period (Fig. 2f–j). The *Hoxb1* and *Hoxb4* grafts did not contribute to the posteriormost end of the embryo, whereas overexpression of more posterior *Hox* genes resulted in labelled cells in the tail bud (compare Fig. 2g, h with i, j). Remarkably, the final distribution along the AP axis of the labelled descendants of the grafts overexpressing *Hoxb1* to *Hoxb9* was strictly collinear (Fig. 2b–e, g–j).

In order to directly compare the effect of different *Hox* genes in the same embryos, we orthotopically grafted at the same streak level two half-primitive streak domains coming from stage-matched embryos overexpressing either *Hoxb1*, *Hoxb4*, *Hoxb7*, *Hoxb9* or control constructs with ZsGreen or DsRed (Fig. 2k, l). We first compared the fate of grafts overexpressing *Hoxb4* labelled with DsRed with grafts overexpressing *Hoxb9* labelled with ZsGreen (Fig. 2l–p). After a short incubation of 6 hours, *Hoxb4* cells were already found in the PSM (Fig. 2m, red), while the *Hoxb9*-expressing cells were still located at the level of the primitive streak (Fig. 2m, green). Analysis of transverse sections demonstrated that cells overexpressing *Hoxb9* were still located in the epiblast layer and that most retained a characteristic epithelial columnar shape (Fig. 2p, green); in contrast, a large fraction

of cells expressing *Hoxb4* had lost their epithelial characteristics and were located in the primitive streak and the ingressed mesoderm (Fig. 2p, red). When embryos were analysed after a 16-hour incubation period, the red-labelled cells were always found to extend several somites anterior to the green-labelled cells (Fig. 2n, o,  $n = 6$ ). When the embryos were incubated an additional 24 hours, only the green-labelled cells were found to extend to the posterior tip of the embryo (Fig. 2o). Similar results were observed with double grafts comparing *Hoxb1*-overexpressing cells labelled with DsRed, and *Hoxb7*-overexpressing cells labelled with ZsGreen (data not shown).

We then grafted orthotopically at an anterior position along the primitive streak a half-primitive streak fragment from a *Hoxb9*-DsRed overexpressing donor and on the same side but slightly more posteriorly, a half-primitive streak from a *Hoxb1*-ZsGreen donor (Fig. 2q, r,  $n = 4$ ). After a 4-hour reincubation period, a large number of descendants of the *Hoxb1*-expressing cells had already



**Figure 2 | *Hox* genes control the timing of ingression of epiblast cells into the primitive streak.** **a–j**, Homotopic and homochronic grafts of fragments of the 80% level of the primitive streak electroporated with pCIZ control vector (**a**, **f**), or *Hoxb1* (**b**, **g**), *Hoxb4* (**c**, **h**), *Hoxb7* (**d**, **i**), *Hoxb9* (**e**, **j**) expression vectors driving internal ribosome entry site 2 (*IRE*S2)-ZsGreen expression following a reincubation period of 16 h (**a–e**) or 40 h (**f–j**). White arrowheads mark the fifth somite level. **k**, Homotopic and homochronic double graft of fragments of the 80% level of the primitive streak from embryos electroporated with *Hoxb4*-*IRE*S2-DsRed and *Hoxb9*-*IRE*S2-ZsGreen, respectively. The grafted embryo is shown just before reincubation (**l**), and 6 h (**m**), 16 h (**n**) and 40 h (**o**) after reincubation. **p**, Section of an embryo grafted as in **k** incubated for 6 h at the level indicated in **m** (white dashed line). **q**, Homotopic and homochronic double graft of fragments of the 80% level of the primitive streak from an embryo electroporated with *Hoxb9*-*IRE*S2-DsRed and of the 70% primitive streak level from an embryo electroporated with *Hoxb1*-*IRE*S2-ZsGreen, respectively. The grafted embryo is shown just before reincubation (**r**), and 4 h (**s**, **v**, **w**), 8 h (**t**) and 28 h (**u**) after reincubation. Hatched lines in **s** indicate the level of transverse sections shown respectively in **v** and **w**. Green and red arrowheads mark the anterior and posterior extension of the descendants of the labelled-grafted cells by each reporter along the AP axis. **a–j**, **l–o**, **r–u**, Ventral views, anterior to the top. ps, primitive streak.

ingressed into the mesoderm (Fig. 2s, v, green); in contrast, most of the *Hoxb9*-expressing cells were still found in the epiblast layer of the primitive streak (Fig. 2s, w, red). After 8 hours, some *Hoxb1*-overexpressing cells were found anterior to the *Hoxb9*-overexpressing cells (Fig. 2t). After a 28-hour reincubation period, the *Hoxb1*-overexpressing cells that had been grafted posteriorly were found to extend more anteriorly than the *Hoxb9*-overexpressing cells (Fig. 2u). In similar experiments in which only control red and green reporters were electroporated, cells grafted anteriorly in the streak were always found to lie more anterior in the somitic series (data not shown, Supplementary Table 1). These experiments indicate that overexpressing more posterior *Hox* genes in epiblast cells can delay the timing at which cells ingress from the epiblast into the primitive streak and the nascent mesoderm. This alters the fate of paraxial mesoderm cells along the AP axis, leading the cells to become more posterior than they normally would.

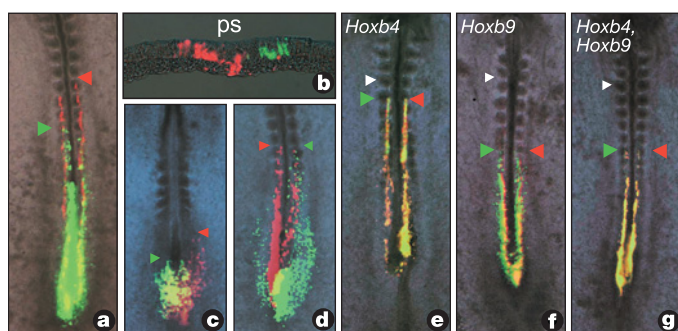
We then developed a procedure to successively electroporate two different constructs labelled with two distinct colours in the anterior primitive streak of chick embryos. This protocol results in essentially distinct cells being transfected with each construct. When we successively electroporated *Hoxb9*-*ZsGreen* and *Hoxb1*-*DsRed*, or *Hoxb9*-*ZsGreen* and a *DsRed*-expressing vector in the primitive streak, the descendants of the cells overexpressing *Hoxb9* were located in a more posterior region than the cells overexpressing *Hoxb1* or the control *DsRed* vector (compare the position of green and red arrowheads in Fig. 3a and data not shown). In all the combinations of *Hox* genes analysed, the descendants of the cells expressing the more 5' gene were always located in a more posterior domain (Supplementary Table 2). Analysis of sections shows that ingression of the cells overexpressing *Hoxb9* (Fig. 3b, green) was delayed compared to cells overexpressing *Hoxb1* (Fig. 3b, red). This effect on cell ingression requires a functional homeodomain, since deleting the helix 3 of the *Hoxb9* homeodomain completely suppressed the delaying effect (Fig. 3c, red) compared with the full length *Hoxb9* overexpression (Fig. 3c, green).

The different roles of *Hox* genes have been proposed to reflect quantitative rather than qualitative variations of expression of these genes<sup>14</sup>. To evaluate the impact of quantitative variations of the same gene on the ingression process, we compared the effect of two *Hoxb9*

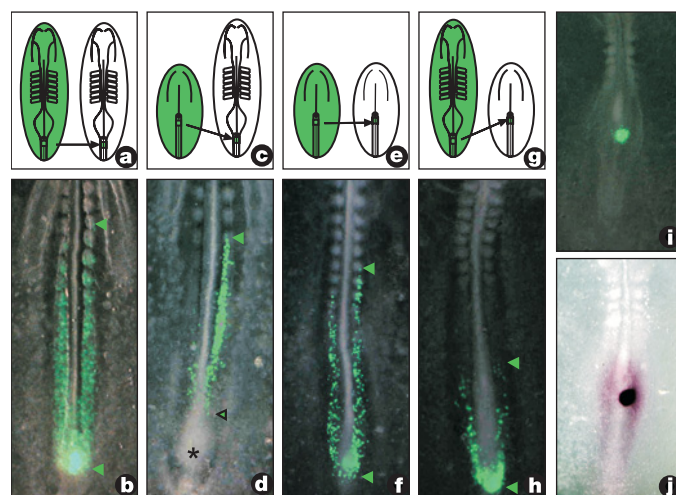
constructs driven either by the strong CAGGS promoter ( $\beta$ -actin promoter and CMV enhancer) and by the weak thymidine kinase promoter (HSV-TK)<sup>15</sup>. Whereas a bright green signal was detected when *pCAGGS-Hoxb9-ZsGreen* was overexpressed in the chick PSM, very weak or no fluorescent signal was detected from the *pTK-Hoxb9-ZsGreen*, indicating that the two promoters drive very different expression levels in this tissue (data not shown). In order to follow the cells overexpressing the *pTK-Hoxb9-ZsGreen* in the embryo, we mixed the plasmid with a *pCAGGS-EGFP* vector. This co-electroporation procedure resulted in the vast majority of cells co-expressing the two vectors. We then electroporated cells sequentially with *pCAGGS-Hoxb9-DsRed* and with *pTK-Hoxb9-ZsGreen* mixed with *pCAGGS-EGFP* as described above. The green and red cells were found to extend to the same anterior level (Fig. 3d), indicating that the low level of *Hoxb9* expression driven from the TK promoter results in the same phenotype as the high level of *Hoxb9* expression from the CAGGS promoter. Similar observations were made with *Hoxb7* or when other weaker promoters like SV40 were compared to the CAGGS promoter<sup>16</sup> (Supplementary Table 2). This indicates that the dose of *Hoxb* gene expression has no significant influence on the final position of the cell along the AP axis.

The results described above also suggest that the 5' *Hox* genes are functionally dominant over the 3' *Hox* genes, consistent with the posterior prevalence of *Hox* genes<sup>17</sup>. To test this directly, we compared the distribution of labelled descendants from 80% level streak grafts overexpressing either *Hoxb4* or *Hoxb9* alone or co-expressing the two genes (Supplementary Table 1). Cells co-expressing *Hoxb4* and *Hoxb9* in somites extend to a similar anterior level to cells expressing *Hoxb9* alone, which lie several somites posterior when compared to cells expressing only *Hoxb4* (compare Fig. 3e–g). A similar effect was observed when co-expressing *Hoxb1* and *Hoxb7* (data not shown), indicating that posterior genes are able to suppress the effect of more anterior *Hox* genes.

These results imply that heterochronic grafts of primitive streak fragments, which normally express different combinations of *Hox*



**Figure 3 | Posterior prevalence of *Hoxb* genes.** **a**, Embryo successively electroporated at the anterior primitive streak level at stage 5HH with *Hoxb9*-*ZsGreen* and *Hoxb1*-*DsRed*, 16 h after incubation. **b**, Transverse section at the same level as the embryo shown in Fig. 2m (hatched line) of an embryo electroporated sequentially with *Hoxb1*-*IRES2*-*DsRed* and *Hoxb9*-*IRES2*-*ZsGreen* fixed 4 h after reincubation. **c**, Embryo successively electroporated with C-terminal-deleted-*Hoxb9*-*DsRed* and *Hoxb9*-*ZsGreen* at 6HH, 5 h after reincubation. **d**, Embryo successively electroporated with *Hoxb9* driven by the CAGGS promoter (in red) and *Hoxb9* driven by the TK promoter (in green) at 5HH, shown 16 h after reincubation. **e–g**, Co-electroporation of the anterior primitive streak region at 5HH with two constructs; **e**, *Hoxb4*-*DsRed* and *ZsGreen* vector; **f**, *DsRed* vector and *Hoxb9*-*ZsGreen*, and **g**, *Hoxb4*-*DsRed* and *Hoxb9*-*ZsGreen*. Green and red arrowheads mark the anterior extension of the descendants of the labelled grafted cells by each reporter along the AP axis. **a**, **c–g**, Ventral views, anterior to the top. ps, primitive streak.



**Figure 4 | Heterochronic grafts of the primitive streak.** **a**, Homotopic and homochronic 80% level primitive streak graft from an EGFP-labelled 6-somite stage embryo. **b**, 20 h after incubation. **c**, Homotopic and heterochronic 80% level primitive streak graft from a stage 6HH EGFP-labelled donor to a 6-somite stage host. **d**, 20 h after incubation. Tail bud is marked by an asterisk. **e**, Homotopic and homochronic 80% level primitive streak graft from an EGFP-labelled 6HH embryo. **f**, 20 h after incubation. **g**, Homotopic and heterochronic 80% level primitive streak graft from a 6-somite stage EGFP-labelled donor to a 6HH host. **h**, 20 h after incubation. **i**, 6 h after incubation. **j**, *Hoxb9* *in situ* hybridization of the embryo shown in **i**. Ventral views, anterior to the top. Green arrowheads mark the anterior and posterior extension of the descendants of the labelled-grafted cells along the AP axis.



genes, should result in predictable outcomes. When a streak fragment originating from a younger host was grafted homotopically into the primitive streak of an older host, labelled cells reached a similar anterior level to homochronic control grafts (Supplementary Table 3; compare Fig. 4a, b and c, d), but no cells were observed in the tail bud (compare Fig. 4b and d). In contrast, when a primitive streak fragment from an older embryo was grafted into a younger host, labelled cells were found to extend from the tail bud to a level located more posteriorly by up to 6 somites than the level of grafted cells in stage-matched controls (Supplementary Table 3; compare Fig. 4e, f and g, h). In grafts from an older primitive streak, no emigration of cells from the grafted tissue was observed following a 6-hour reincubation period (Fig. 4i), consistent with the strong expression of posterior *Hox* genes like *Hoxb9* which was retained in the grafted cells (Fig. 4j). These findings are consistent with earlier observations in the mouse showing that heterochronic grafts of older tail bud into earlier mouse embryos result in grafted cells being located more posteriorly along the AP axis than expected<sup>18</sup>. Even though there might be more differences between an old and a younger graft from the same streak level than the mere expression of *Hox* genes, these observations support the idea that *Hox* genes expressed in the grafted territories at these stages control the ingression of somite derivatives from the graft.

Our data therefore suggest that at a defined time point, ingression of epiblast cells is controlled by the most 5' *Hox* genes expressed by these cells. When the next more 5' paralogous *Hox* gene becomes activated in a subpopulation of these epiblast cells, these cells will acquire slightly different migratory properties, and their ingression will be slightly delayed. Expression of this *Hox* gene rapidly spreads to the surrounding cells that now exhibit similar migratory properties until expression of the next 5' *Hox* gene begins in scattered epiblast cells, altering their migratory properties once again. Overall, this process would ensure that ingressing cells expressing *Hox* genes from consecutive paralogous groups will sort out from each other along the AP axis. This ordered ingression of epiblast cells in the streak probably constitutes an initial step for the establishment of the nested *Hox* expression domains of the mesoderm (that is, the spatial co-linearity).

## METHODS

**Embryos and nomenclature.** Fertilized chick eggs were obtained from commercial sources. Eggs were incubated at 38 °C in a humidified incubator, and embryos were staged according to Hamburger and Hamilton (HH)<sup>19</sup> and by counting somites.

**RNA *in situ* hybridization and probes.** Whole mount RNA *in situ* hybridizations were carried out as described<sup>20</sup>. Chicken *Hoxb1*, *Hoxb2*, *Hoxb3*, *Hoxb4*, *Hoxb5*, *Hoxb6*, *Hoxb7*, *Hoxb8* and *Hoxb9* probes have been described<sup>21</sup>. Two-colour *in situ* hybridization is described in Supplementary Information. Some embryos were embedded for cryosection and sliced at 14 µm.

**Electroporation and fate mapping of the epiblast.** Embryos ranging from stage 3HH to stage 7HH were prepared for EC culture<sup>22</sup>. *In vitro* electroporations were carried out as described in Supplementary Information. Small fragments of electroporated epiblast with unlabelled underlying mesoderm and endoderm were excised and grafted into the same position of stage-matched unlabelled host embryos<sup>23</sup>. After reincubation, the position of the fluorescent cells in the embryos was scored.

**Hox expression constructs.** *Hox* expression vectors were generated in expression vectors driven by different promoters including *pCIZ* and *pCIRX*, which are originally derived from the *pCAGGS* expression vector<sup>24</sup>.

More detailed experimental procedures are available in Supplementary Information.

Received 24 January; accepted 2 May 2006.

Published online 7 June 2006.

1. Kmita, M. & Duboule, D. Organizing axes in time and space; 25 years of colinear tinkering. *Science* **301**, 331–333 (2003).
2. Krumlauf, R. *Hox* genes in vertebrate development. *Cell* **78**, 191–201 (1994).

3. Dolle, P., Izpisua-Belmonte, J. C., Falkenstein, H., Renucci, A. & Duboule, D. Coordinate expression of the murine *Hox-5* complex homeobox-containing genes during limb pattern formation. *Nature* **342**, 767–772 (1989).
4. Gaunt, S. J., Sharpe, P. T. & Duboule, D. Spatially restricted domains of homeo-gene transcripts in mouse embryos: relation to a segmented body plan. *Development* **104**, 169–179 (1988).
5. Graham, A., Papalopulu, N. & Krumlauf, R. The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell* **57**, 367–378 (1989).
6. Kessel, M. & Gruss, P. Homeotic transformations of murine vertebrae and concomitant alteration of *Hox* codes induced by retinoic acid. *Cell* **67**, 89–104 (1991).
7. Wellik, D. M. & Capecchi, M. R. *Hox10* and *Hox11* genes are required to globally pattern the mammalian skeleton. *Science* **301**, 363–367 (2003).
8. Deschamps, J. & van Nes, J. Developmental regulation of the *Hox* genes during axial morphogenesis in the mouse. *Development* **132**, 2931–2942 (2005).
9. Wacker, S. A., Jansen, H. J., McNulty, C. L., Houtzager, E. & Durston, A. J. Timed interactions between the *Hox* expressing non-organiser mesoderm and the Spemann organiser generate positional information during vertebrate gastrulation. *Dev. Biol.* **268**, 207–219 (2004).
10. Forlani, S., Lawson, K. A. & Deschamps, J. Acquisition of *Hox* codes during gastrulation and axial elongation in the mouse embryo. *Development* **130**, 3807–3819 (2003).
11. Stern, C. D. in *Gastrulation: From Cells to Embryos* (ed. Stern, C. D.) 219–232 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2004).
12. Garcia-Martinez, V., Alvarez, I. S. & Schoenwolf, G. C. Locations of the ectodermal and nonectodermal subdivisions of the epiblast at stages 3 and 4 of avian gastrulation and neurulation. *J. Exp. Zool.* **267**, 431–446 (1993).
13. Nicolet, G. Avian gastrulation. *Adv. Morphog.* **9**, 231–262 (1971).
14. Greer, J. M., Puetz, J., Thomas, K. R. & Capecchi, M. R. Maintenance of functional equivalence during paralogous *Hox* gene evolution. *Nature* **403**, 661–665 (2000).
15. Stewart, C. L., Schuetze, S., Vanek, M. & Wagner, E. F. Expression of retroviral vectors in transgenic mice obtained by embryo infection. *EMBO J.* **6**, 383–388 (1987).
16. Stamatakis, D., Ulloa, F., Tsoni, S. V., Mynett, A. & Briscoe, J. A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube. *Genes Dev.* **19**, 626–641 (2005).
17. Duboule, D. & Morata, G. Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet.* **10**, 358–364 (1994).
18. Tam, P. P. & Tan, S. S. The somitogenetic potential of cells in the primitive streak and the tail bud of the organogenesis-stage mouse embryo. *Development* **115**, 703–715 (1992).
19. Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *Dev. Dyn.* **195**, 231–272 (1992).
20. Henrique, D. et al. Expression of a Delta homologue in prospective neurons in the chick. *Nature* **375**, 787–790 (1995).
21. Bel-Vialar, S., Itasaki, N. & Krumlauf, R. Initiating *Hox* gene expression: in the early chick neural tube differential sensitivity to FGF and RA signaling subdivides the *HoxB* genes in two distinct groups. *Development* **129**, 5103–5115 (2002).
22. Chapman, S. C., Collignon, J., Schoenwolf, G. C. & Lumsden, A. Improved method for chick whole-embryo culture using a filter paper carrier. *Dev. Dyn.* **220**, 284–289 (2001).
23. Yang, X., Dormann, D., Munsterberg, A. E. & Weijer, C. J. Cell movement patterns during gastrulation in the chick are controlled by positive and negative chemotaxis mediated by FGF4 and FGF8. *Dev. Cell* **3**, 425–437 (2002).
24. Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193–199 (1991).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Abmayr, R. Krumlauf and D. Wellik for comments on the manuscript; X.-S. Yang and C. Weijer for sharing their expertise in fate mapping and for discussions; members of the Pourqu  laboratory for sharing reagents and for discussions; C. Tomomori-Sato and S. Sato of the Conaway laboratory for discussions; and S. Esteban for artwork. This work was supported by the Stowers Institute for Medical Research and the NIH. O.P. is a Howard Hughes Medical Institute Investigator.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to O.P. ([olp@stowers-institute.org](mailto:olp@stowers-institute.org)).



## LETTERS

# Norm-based face encoding by single neurons in the monkey inferotemporal cortex

David A. Leopold<sup>1†</sup>, Igor V. Bondar<sup>1†</sup> & Martin A. Giese<sup>2</sup>

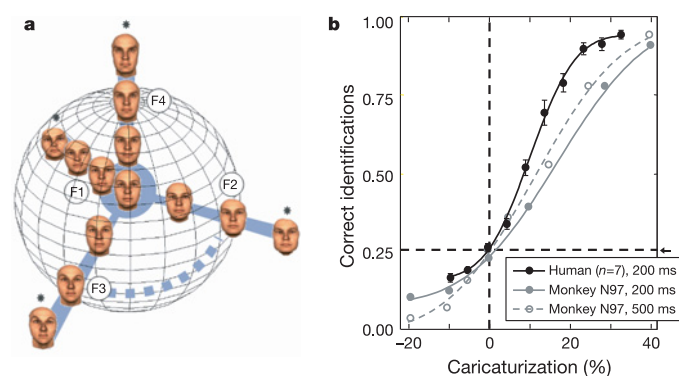
The rich and immediate perception of a familiar face, including its identity, expression and even intent, is one of the most impressive shared faculties of human and non-human primate brains. Many visually responsive neurons in the inferotemporal cortex of macaque monkeys respond selectively to faces<sup>1–4</sup>, sometimes to only one or a few individuals<sup>5–7</sup>, while showing little sensitivity to scale and other details of the retinal image<sup>8,9</sup>. Here we show that face-responsive neurons in the macaque monkey anterior inferotemporal cortex are tuned to a fundamental dimension of face perception. Using a norm-based caricaturization framework previously developed for human psychophysics<sup>10–12</sup>, we varied the identity information present in photo-realistic human faces<sup>13</sup>, and found that neurons of the anterior inferotemporal cortex were most often tuned around the average, identity-ambiguous face. These observations are consistent with face-selective responses in this area being shaped by a figural comparison, reflecting structural differences between an incoming face and an internal reference or norm. As such, these findings link the tuning of neurons in the inferotemporal cortex to psychological models of face identity perception.

Previous work showing that face recognition benefits from caricaturization—that is, the exaggeration of distinguishing features<sup>10,14–16</sup>—has led to the conceptualization of a multidimensional ‘face space’ (see cartoon in Fig. 1a), within which each point represents an individual face and the average of all faces resides at the centre<sup>17</sup>. We focused on stimuli along four face-space trajectories branching away from the average face (sometimes termed the ‘norm’ or ‘prototype’ face) towards increasing identity levels. This radial axis, sometimes termed an axis of caricaturization or identity trajectory, has been studied extensively, and is thought to play an important role in face perception<sup>10–12,18,19</sup>. Also shown are faces separated tangentially, morphed directly between faces of different identities. The construction of photo-realistic faces along radial and tangential trajectories in a face space were generated with a ‘morphable face model’, which interpolates between three-dimensional laser scans of human heads<sup>13</sup>. A recent study using these stimuli found that equal increments along the caricaturization axes in a face space defined by such a morphable model translated directly to equal perceptual increments<sup>12</sup>. These data are shown for humans in Fig. 1b, and are compared directly to similar data from one of the monkeys used in this study (N97). This monkey recognized human faces in a manner similar to human subjects (albeit with a slightly diminished performance), and experienced a face-identity after-effect in the same way as human subjects<sup>20</sup> (see also Supplementary Fig. 2).

To test whether the psychological representation of the caricaturization axes was reflected in the firing of single neurons in the

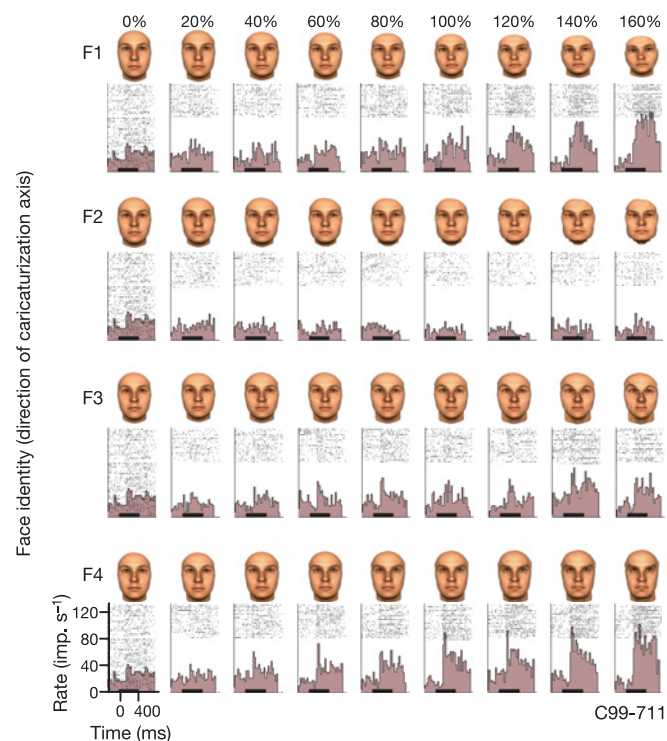
inferotemporal cortex, we performed extracellular microelectrode recordings in two monkeys (see Methods) that were presented with a large number of faces corresponding to different positions in face space. The faces were presented in a pseudorandom sequence. The microelectrodes were situated in the anterior inferotemporal cortex (AIT) region (see Supplementary Fig. 3), an area known to contain neurons that respond selectively to face and other complex patterns. Each day, the monkey was presented with faces from four different axes of caricaturization, corresponding to four different individuals. Different sets of faces were used for the two monkeys. One of the animals, C99, was additionally tested with faces modulated along tangential directions of face space (see dotted line in Fig. 1a), corresponding to morphs between different face identities. A total of 250 face-responsive units were monitored, with 209 tested with a complete set of faces for our evaluation. Despite several differences in the training and testing, and in the electrophysiological procedures (see Methods), the results were similar in both animals.

The main finding was a striking tendency for neurons to show tuning that appeared centred about the average face. For a large proportion of neurons, this entailed monotonic trends in firing for



**Figure 1 | Face space and behavioural data.** **a**, Each identity trajectory (F1–F4) intersects in the centre, at the average face (0% identity). Identity levels change from the average through to the full identity (100%) faces, to the caricatures (160% shown, asterisks). Other face identities in this study (for example, between F2 and F3, dotted blue line), lay on the tangential trajectories. **b**, Behavioural performance of monkey N97 in the recognition task, shown for two different face presentation durations and compared to seven humans (mean  $\pm$  s.e.m.) performing the same task (with 500-ms presentations). Chance performance for correctly recognizing a face out of four learned identities was 0.25 (horizontal dashed line). Vertical dashed line corresponds to zero identity level.

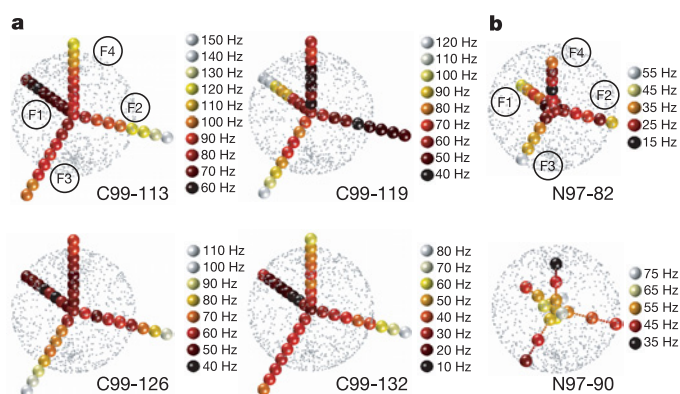
<sup>1</sup>Max Planck Institut für biologische Kybernetik, Spemannstrasse 38, 72076 Tübingen, Germany. <sup>2</sup>Laboratory for Action Representation and Learning, Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, University Clinic Tübingen, Schaffhausenstr. 113, D-72072 Tübingen, Germany. <sup>†</sup>Present addresses: Unit on Cognitive Neurophysiology and Imaging, Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, 49 Convent Drive, Bldg. 49, MSC 4400, Bethesda, Maryland 20892, USA (D.A.L.); Institute of Higher Nervous Activity and Neurophysiology, Russian Academy of Sciences, Butlerova Street 5a, 117485 Moscow, Russia (I.V.B.).



**Figure 2 | Responses of a single neuron (C99-117) to faces along axes of caricaturization.** Each panel corresponds to the rasters and peristimulus time histogram in response to a 400-ms presentation of the face stimulus (dark bar); see panel at bottom left for axis information (imp., impulse). Rows represent each radial axis in face space for faces F1–F4, from the average face (left, identical in all cases) through to the full identity face (100%), to the caricatures (right).

increasing levels of caricaturization, for two or more of the face trajectories. The example in Fig. 2 shows a typical result, for four faces modulated at nine caricaturization values ranging from 0% (average face) through to 100% (full identity) and up to 160% (caricature). This neuron, like many we observed, showed a roughly linear increase in the discharge of action potentials as a function of face identity. It was able to reliably discriminate between extremely subtle differences in the successive faces along the radial dimension, even for identity levels well above the recognition threshold suggested by the behavioural data. Additional examples are represented in a more succinct form in Fig. 3, as coloured spheres emanating from a central sphere (the average face), with the geometry illustrated in Fig. 1a. More traditional tuning curves for a large number of individual neuron examples for both monkeys are provided in Supplementary Figs 4 and 5, further demonstrating the apparently special role of the average face in the firing of AIT neurons. Interestingly, neurons with extreme near-linear tuning appeared to be clustered<sup>21</sup>, and were typically encountered within the same recording session. The examples in Fig. 3a are the responses of four neurons collected from four different electrodes within an estimated 250  $\mu\text{m}$  of each other in the same session.

Across the population, each of the four face identities elicited similar, near-linear changes in response magnitude as a function of caricaturization level for both monkeys (Fig. 4a). We next asked for how many of the four identities, on average, did the neurons show monotonic trends with the identity level, and whether or not these trends had a tendency to go in the same direction. To address this question, we first rank-ordered the responses to the four identities for each neuron, and then computed the average normalized rank-ordered profile (Fig. 4b). The results show that, on average, between two and three of the four faces showed significant monotonic activity changes in the same direction with increasing identity levels. Note

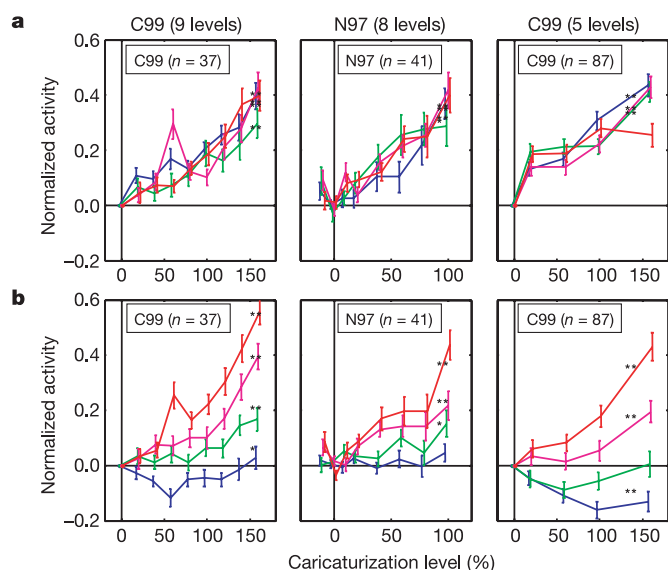


**Figure 3 | Response magnitudes for different caricaturization levels of all four faces.** Each tetrahedron corresponds to the responses of a single AIT neuron for monkey C99 (a) and N97 (b). Labels below (for example, C99-113) are identifiers for each neuron from monkey C99 and N97. Each sphere comprising the tetrahedra is coloured according to the neuron's mean spike rate following 8–12 presentations of the face at that caricaturization level. The four neurons in a were collected on nearby electrodes during the same session.

that in the case of monkey N97, which was presented a negative caricaturization level (–10%), there is a response minimum at the average face, revealing that both positive and negative deviations from the average face in either direction leads, at the population level, to increased responses. Systematic trends in the data were not observed when the data sets were randomly reordered (see Supplementary Fig. 6).

Finally, in monkey C99, we presented faces along tangential directions in face space, corresponding to morphs between different facial identities. Several representative examples of responses to both radial and tangential morphs are shown in Fig. 5a, in the same representation as in Fig. 3. These examples illustrate the diversity of tuning we observed along the tangential direction, and further show the tendency towards monotonic trends in the radial (caricaturization) axes. In addition, traditional tuning curves along the tangential direction, between all combinations of face identities, are shown for four examples in Supplementary Fig. 7, which demonstrates the high variability in the population. The overall tangential tuning, averaged for the entire population of 87 neurons, is shown in Fig. 5b. Note that the data panels are sorted in order, from the face that gave the lowest response (R1) to that which gave the highest response (R4). The population responses reveal that the tuning was, on average, also largely monotonic between faces. All these monotonic trends were significant, and were no longer present after reshuffling (Supplementary Fig. 8).

These findings, taken together, demonstrate that the axis of caricaturization, known to be a meaningful perceptual dimension in face recognition, is also a primary determinant of the firing of face-selective neurons in the inferotemporal cortex. The special role of the average face in shaping the tuning of neurons in this area may reflect the brain's plasticity in capturing the natural variation in facial characteristics. Considerable electrophysiological evidence suggests that the tuning in area AIT can be shaped by relevant statistics of the visual input<sup>22–24</sup>, and it is therefore likely that the tuning we observed was a result of exposure to this particular stimulus set. It is also difficult to disentangle the tuning of these cells from factors such as stimulus familiarity<sup>25</sup> or response certainty, though the similar results obtained for the two monkeys argues against this playing a major role in the neural responses. Furthermore, neural responses to the identity-ambiguous average face were only minimally correlated with identity reported from trial to trial, reaching significance in only 15/84 neurons where such analysis was possible (monkey N97, one-way analysis of variance,  $P < 0.05$ ). It therefore appears that,



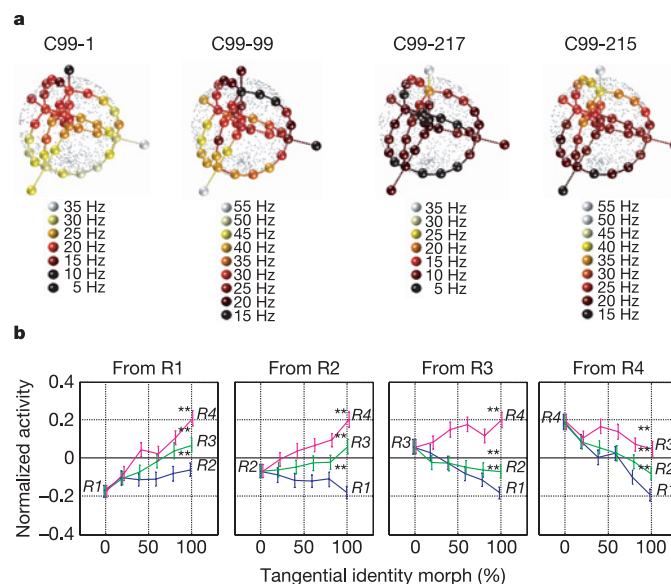
**Figure 4 | Population responses as a function of the caricaturization level.** **a**, Population responses for each monkey (C99 participated in some sessions with five levels and some with nine) for all four of the faces used, which were different for the two animals. **b**, Faces were rank-ordered, from the one that elicited maximum response to the one that elicited minimum response for each neuron, and then averaged for each monkey. Normalized responses were averaged separately for each rank and caricaturization level. Asterisks indicate a significant deviation from a constant trend (\* $P = 0.05$ ; \*\* $P = 0.01$ ; error bars, s.e.m.; see Methods and Supplementary Information for details).

once tuning for faces in AIT has been established, neural responses are determined primarily by the physical structure of each face stimulus. Interestingly, a recent electrophysiological study demonstrated that neurons tuned for different shapes tend to show their maximum responses for the extreme values of those shapes<sup>26</sup>, suggesting that the brain may exploit a norm-based approach for a wide range of visual stimuli and tasks.

We conclude that the face-selective responses reported in the present study may represent the outcome of an internal, comparative process, perhaps expressed over multiple stages of visual processing that may allow the brain to decipher subtle differences in the shape and configuration of facial features that form the basis of face recognition. In this context, it is interesting to consider that norm-based processing may contribute to the visual system's robustness to stimulus transformations encountered in normal vision. Common transformations such as changes in scale and viewing angle, facial expression or even age need not be applied and learned for each facial identity, but might instead be applied to the normalizing pattern to which incoming faces are compared, saving both the time required to learn invariant representations of each face, as well as the corresponding storage capacity in the brain. Norm-based mechanisms, having adapted to our precise needs in face recognition, may also help explain why our face recognition is so immediate and effortless, and why at the same time we have such poor intuitive insight into how we accomplish this task.

## METHODS

**Subjects and task.** Two monkeys (*Macaca mulatta*), N97 and C99, were used in this study, each trained on a different set of faces. Although both animals had viewed the faces extensively (that is, had been trained on them for several weeks), by the time of testing only monkey N97 (male) was required to explicitly learn and respond to the face identity. Specifically, this monkey was trained in a four alternative forced-choice procedure to report, with two bidirectional levers, which of the four identities had been presented. This monkey's performance reached near perfect with roughly 40% face identity, which was slightly worse than that of humans tested with the same stimuli<sup>12</sup> (see Fig. 1b, and Supplementary



**Figure 5 | Responses to radial and tangential morphs.** **a**, Tetrahedron representation as in Fig. 3, but now with tangential directions included. The four tetrahedrons correspond to four examples from C99 collected in different recording sessions. **b**, Rank-ordered population responses along the tangential morph trajectories (for details see Methods). On each panel, three curves trace each of the trajectories away from a given face, from that yielding the minimum response (R1) to that leading the maximum response (R4) (conventions same as in Fig. 4).

Fig. 2). In contrast, monkey C99 (male) was not trained to respond to the identity of the test faces.

Daily testing proceeded as follows. Upon hearing a brief tone, the monkey was required to fixate a small point that appeared in the centre of the screen. After 500 ms, the test face appeared and remained on the screen for 400 ms before disappearing. The fixation requirements were minimal, with the trial aborting only if the animal's gaze was not directed towards the face at all. Monkeys received apple juice as a reward, with between 800 and 1,200 trials collected in a typical recording session. Monkey C99 was rewarded as long as it fixated throughout the entire trial, whereas monkey N97 was rewarded on a variable reward schedule for correct responses (for more details, see ref. 20).

The stimuli were generated using a morphable face model<sup>13</sup>, which was based on the Face Database of the Max Planck Institute for Biological Cybernetics in Tübingen, Germany (<http://faces.kyb.tuebingen.mpg.de/index.php>)<sup>13</sup>. Each day, the spiking of single units was monitored as the monkey was presented with between 16 and 41 different faces in pseudorandom order, at least 8 times each and often more than 20 times each. For additional details on stimuli, electrophysiological procedures and surgical procedures, see Supplementary Information.

**Analysis.** Data were analysed using custom software written in Matlab (Mathworks Inc.). Tuning curves were computed using the mean discharge rate of well-isolated single neurons, in a time interval from 120 to 370 ms following stimulus onset. For the population analysis, the tuning curves for each neuron were first normalized by subtracting the response to the average face, and then dividing by the maximum response elicited by any face. For further details, see Supplementary Information.

Received 7 February; accepted 12 May 2006.

Published online 5 July 2006.

- Perrett, D. I., Rolls, E. T. & Caan, W. Visual neurones responsive to faces in the monkey temporal cortex. *Exp. Brain Res.* **47**, 329–342 (1982).
- Gross, C. G. How inferior temporal cortex became a visual area. *Cereb. Cortex* **4**, 455–469 (1994).
- Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B. & Tootell, R. B. Faces and objects in macaque cerebral cortex. *Nature Neurosci.* **6**, 989–995 (2003).
- Logothetis, N. K. & Sheinberg, D. L. Visual object recognition. *Annu. Rev. Neurosci.* **19**, 577–621 (1996).
- Hasselmo, M. E., Rolls, E. T. & Baylis, G. C. The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav. Brain Res.* **32**, 203–218 (1989).
- Sugase, Y., Yamane, S., Ueno, S. & Kawano, K. Global and fine information



- coded by single neurons in the temporal visual cortex. *Nature* **400**, 869–873 (1999).
7. Perrett, D. I. *et al.* Neurons responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception. *Hum. Neurobiol.* **3**, 197–208 (1984).
  8. Rolls, E. T. & Baylis, G. C. Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Exp. Brain Res.* **65**, 38–48 (1986).
  9. Tovee, M. J., Rolls, E. T. & Azzopardi, P. Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. *J. Neurophysiol.* **72**, 1049–1060 (1994).
  10. Rhodes, G., Brennan, S. & Carey, S. Identification and ratings of caricatures: implications for mental representations of faces. *Cogn. Psychol.* **19**, 473–497 (1987).
  11. Valentine, T. & Bruce, V. The effects of distinctiveness in recognising and classifying faces. *Perception* **15**, 525–535 (1986).
  12. Leopold, D. A., O'Toole, A. J., Vetter, T. & Blanz, V. Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neurosci.* **4**, 89–94 (2001).
  13. Blanz, V. & Vetter, T. in *Proc. SIGGRAPH'99* 187–194 (Association for Computing Machinery, New York, Computer Society Press, 1999).
  14. Lee, K., Byatt, G. & Rhodes, G. Caricature effects, distinctiveness, and identification: testing the face-space framework. *Psychol. Sci.* **11**, 379–385 (2000).
  15. Lewis, M. B. & Johnston, R. A. Understanding caricatures of faces. *Q. J. Exp. Psychol. A* **51**, 321–346 (1998).
  16. Mauro, R. & Kubovy, M. Caricature and face recognition. *Mem. Cogn.* **20**, 433–440 (1992).
  17. Valentine, T. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol. A* **43**, 161–204 (1991).
  18. Rhodes, G., *et al.* in *Fitting the Mind to the World: Adaptation and Aftereffects in High-Level Vision* (eds Clifford, C. W. & Rhodes, G.) 213–240 (Oxford Univ. Press, Oxford, UK, 2005).
  19. Leopold, D. A., Rhodes, G., Mueller, K. M. & Jeffery, L. Dynamics of the face identity aftereffect. *Proc. R. Soc. Lond. B* **272**, 897–904 (2005).
  20. Leopold, D. A. & Bondar, I. V. in *Fitting the Mind to the World: Adaptation and Aftereffects in High-Level Vision* (eds Clifford, C. W. G. & Rhodes, G.) 189–212 (Oxford Univ. Press, Oxford, UK, 2005).
  21. Tsao, D. Y., Freiwald, W. A., Tootell, R. B. & Livingstone, M. S. A cortical region consisting entirely of face-selective cells. *Science* **311**, 670–674 (2006).
  22. Sigala, N. & Logothetis, N. K. Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* **415**, 318–320 (2002).
  23. Baker, C. I., Behrmann, M. & Olson, C. R. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neurosci.* **5**, 1210–1216 (2002).
  24. Kobatake, E., Wang, G. & Tanaka, K. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J. Neurophysiol.* **80**, 324–330 (1998).
  25. Persson, J., Habib, R. & Nyberg, L. Decreased activity in inferotemporal cortex during explicit memory: dissociating priming, novelty detection, and recognition. *Neuroreport* **13**, 2181–2185 (2002).
  26. Kayaert, G., Biederman, I., Op de Beeck, H. P. & Vogels, R. Tuning for shape dimensions in macaque inferior temporal cortex. *Eur. J. Neurosci.* **22**, 212–224 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to G. Rhodes, N. Kriegeskorte and N. K. Logothetis for critical comments on the manuscript, and to A. Maier, M. Wilke and K.-M. Mueller for discussions. We especially thank C. Wallraven, B. Knappmeyer and H. H. Bülhoff for help with the generation of the face stimuli, and for providing data from psychophysical experiments (in humans) that were used for selecting the perceptually most dissimilar faces serving as the stimuli for C99. This work was executed at the Max Planck Institute for Biological Cybernetics, and was supported by a Max Planck Society grant to N. K. Logothetis, as well as grants from the Volkswagen Foundation, DFG and HFSP (to M.A.G.).

**Author Contributions** D.A.L. and I.V.B. originated the experiment and performed the neurophysiological recordings. All authors performed data analysis and discussed the results. D.A.L. and M.A.G. wrote the paper.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.A.L. ([leopoldd@mail.nih.gov](mailto:leopoldd@mail.nih.gov)).

## LETTERS

# Tumorigenic transformation by CPI-17 through inhibition of a merlin phosphatase

Hongchuan Jin<sup>1,2†</sup>, Tobias Sperka<sup>1,2</sup>, Peter Herrlich<sup>1,2</sup> & Helen Morrison<sup>1,2</sup>

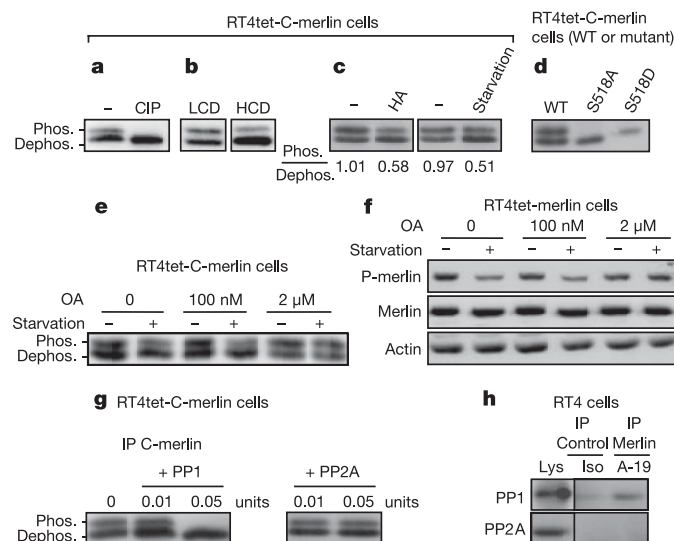
The tumour suppressor protein merlin (encoded by the neurofibromatosis type 2 gene *NF2*) is an important regulator of proliferation in many cell and tissue types<sup>1–4</sup>. Merlin is activated by dephosphorylation at serine 518 (S518), which occurs on serum withdrawal or on cell–cell or cell–matrix contact<sup>5,6</sup>. However, the relevant phosphatase that activates merlin's tumour suppressor function is unknown. Here we identify this enzyme as the myosin phosphatase (MYPT-1–PP1 $\delta$ ). The cellular MYPT-1–PP1 $\delta$ -specific inhibitor CPI-17 causes a loss of merlin function characterized by merlin phosphorylation, Ras activation and transformation. Constitutively active merlin (S518A) reverses CPI-17-induced transformation, showing that merlin is the decisive substrate of MYPT-1–PP1 $\delta$  in tumour suppression. In addition we show that CPI-17 levels are raised in several human tumour cell lines and that the downregulation of CPI-17 induces merlin dephosphorylation, inhibits Ras activation and abolishes the transformed phenotype. MYPT-1–PP1 $\delta$  and its substrate merlin are part of a previously undescribed tumour suppressor cascade that can be hindered in two ways, by mutation of the *NF2* gene and by upregulation of the oncoprotein CPI-17.

We tested whether carboxy-terminal merlin (amino-acid residues 299–595; C-merlin) carrying the decisive phosphorylation site S518 (ref. 6) could be dephosphorylated and could therefore be used as a substrate *in vitro*. C-merlin expressed in rat schwannoma RT4 cells under a doxycycline-inducible promoter (RT4tet-C-merlin)<sup>5</sup> migrated as a distinct doublet. The species with decreased mobility was phosphorylated and converted by treatment with calf intestinal phosphatase to a faster-migrating dephosphorylated band (Fig. 1a) with the same mobility as the non-phosphorylatable S518A mutant protein (in which S518 has been replaced with alanine, in contrast with the pseudo-phosphorylated S518D mutant protein containing an aspartic residue; Fig. 1d). The dephosphorylation of C-merlin was regulated by the same cues as that of full-length merlin<sup>5</sup>, for example, high cell density (Fig. 1b), the addition of high-molecular-mass hyaluronan (HA)<sup>5</sup> or serum withdrawal (as shown in Fig. 1c by the change in the ratio of the upper to the lower band). Because dephosphorylation took less than 5 min after hyaluronan addition or serum withdrawal, we assumed that the activation of a phosphatase was involved rather than a change in stability of the phosphorylated band.

To search for the phosphatase that activates merlin's tumour suppressor function, we tested several phosphatase inhibitors for their effects on merlin dephosphorylation. Merlin dephosphorylation induced by hyaluronan addition (not shown) or serum withdrawal was inhibited by high but not low concentration of okadaic acid (C-merlin in Fig. 1e; full-length merlin in Fig. 1f; merlin phosphorylation status was detected by a phospho-S518 antibody because full-length merlin resolves less well into two bands). Okadaic acid inhibits the protein serine/threonine phosphatase PP2A at

nanomolar concentrations, whereas PP1 is blocked at 10–100-fold higher concentrations<sup>7</sup>. The inhibition characteristics indicated that a PP1-type enzyme might be responsible for merlin dephosphorylation. In an *in vitro* assay, immunoprecipitated partly phosphorylated C-merlin was indeed completely dephosphorylated by 0.05 units of PP1 but not by 0.05 units of PP2A (Fig. 1g).

Although the catalytic process is fast and transient, enzymes sometimes interact with their substrates relatively stably<sup>7,8</sup>. Indeed, PP1 was co-precipitated with endogenous merlin (Fig. 1h), whereas PP2A could not be detected. PP1 proteins accept many substrates, with specificity being accomplished by specific targeting subunits. We considered it likely that the co-precipitation of PP1 with merlin was mediated by one of these targeting subunits. A prominent band



**Figure 1 | PP1 is the catalytic subunit of merlin phosphatase.**

**a**, Immunoprecipitated C-merlin was incubated with or without calf intestinal phosphatase (CIP), resolved by SDS–PAGE and detected by immunoblotting (C-18, merlin antibody). **b**, Immunoblots of C-merlin from cells growing at low cell density (LCD) or high cell density (HCD). **c**, Immunoblots of C-merlin from cells subjected to serum withdrawal (5 min) or incubation with hyaluronan (HA; 100 µg ml<sup>−1</sup> for 5 min). **d**, Gel migration of mutant (S518A, S518D) C-merlin relative to that of the wild type (WT). **e**, **f**, Serum-withdrawal-induced dephosphorylation of C-merlin (**e**) and of full-length merlin probed with phospho-merlin-specific antibodies (**f**) is inhibited by pretreatment of cells with okadaic acid (OA). **g**, *In vitro* dephosphorylation of immunoprecipitated (IP) C-merlin by recombinant PP1 and PP2A. **h**, Immunoprecipitation (IP) of endogenous merlin (merlin antibody A-19 compared to an isotope control antibody iso). Co-precipitated proteins detected with pan PP1 and PP2A antibodies.

<sup>1</sup>Leibniz Institute of Age Research-Fritz-Lipmann-Institute, Beutenbergstrasse 11, 07745 Jena, Germany. <sup>2</sup>Forschungszentrum Karlsruhe, Institute of Toxicology and Genetics, PO Box 3640, 76021 Karlsruhe, Germany. <sup>†</sup>Present address: Cancer Center, Department of Clinical Oncology, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong SAR, China.

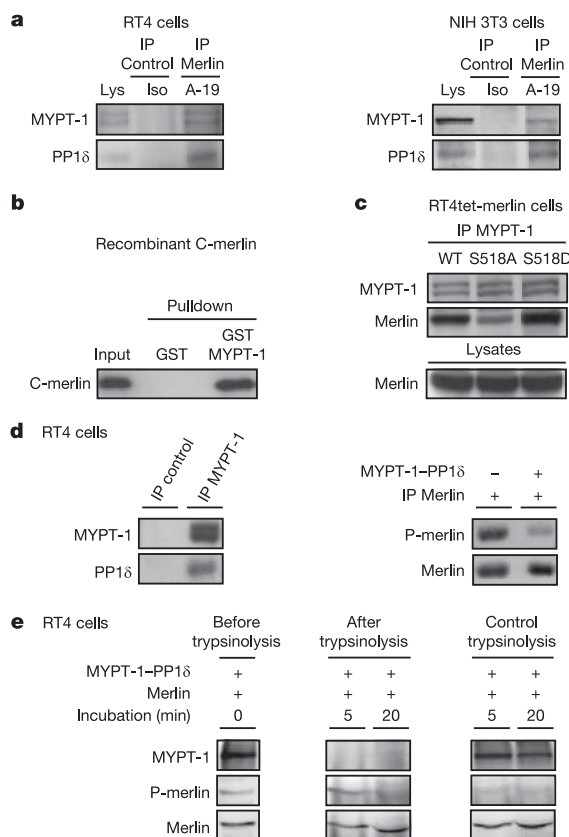
co-immunoprecipitating with C-merlin was identified by mass spectrometry analysis as myosin (myosin heavy chain; Supplementary Fig. S1a). Although a direct interaction between merlin and myosin is possible, we considered a complex with other partners to be more likely. Myosin is known to bind to a phosphatase-targeting subunit termed MYPT-1 (myosin phosphatase targeting subunit 1)<sup>9,10</sup>. Moreover, moesin, a protein with significant sequence homology with merlin, had previously been shown to associate with MYPT-1 (ref. 8). Taken together, these results indicated that MYPT-1 might be the merlin-targeting subunit for PP1. The Coomassie-stained gel also revealed a band at a molecular mass of about 130 kDa, which by immunoblot was identified as MYPT-1 (Supplementary Fig. S1a, b).

MYPT-1 prefers PP1 $\delta$  as the catalytic subunit<sup>10</sup>. Indeed, PP1 $\delta$  and MYPT-1 were co-precipitated with endogenous merlin from lysates of both RT4 and NIH 3T3 cells (Fig. 2a). A bacterially expressed glutathione *S*-transferase (GST) fusion of the C-terminal half of MYPT-1 pulled down recombinant C-merlin *in vitro*, proving a direct interaction between MYPT-1 and merlin (Fig. 2b). Although bound to the non-phosphorylated bacterially expressed C-merlin, MYPT-1 preferred the phosphorylated form of merlin, because full-length S518D or wild type merlin compared to S518A merlin co-precipitated with MYPT-1 (Fig. 2c and Supplementary Fig. S2). In addition to the direct interaction of MYPT-1 with merlin we showed that an immunoprecipitated MYPT-1–PP1 $\delta$  complex accepts phospho-merlin as a substrate *in vitro* (Fig. 2d). To determine whether MYPT-1 targets PP1 $\delta$  to merlin or increases the specific activity of

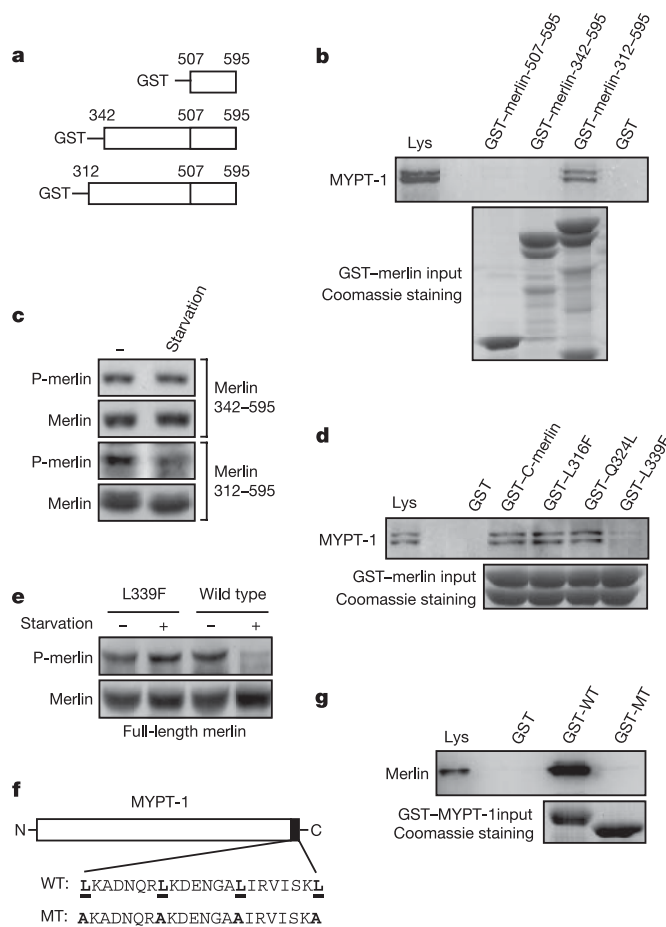
PP1 $\delta$  towards merlin we performed trypsinolysis<sup>12</sup>. However, the degradation of MYPT-1 by trypsin, which releases a fully active PP1 $\delta$ , decreased merlin dephosphorylation, indicating that MYPT-1 acts as a substrate specifier (Fig. 2e).

We determined by GST pull-downs (constructs in Fig. 3a) the interacting domains of both merlin and MYPT-1. Only C-merlin-312–595, but not C-merlin-507–595 or C-merlin-342–595, precipitated MYPT-1 from lysates of RT4 cells (Fig. 3b). The interacting domain (residues 312–341) seems to be the amino-terminal part of the  $\alpha$ -helical coiled-coil region of merlin<sup>13</sup>, which is coarsely consistent with the MYPT-1-interacting region in moesin<sup>8</sup>. MYPT-1 binding was essential for merlin dephosphorylation *in vivo* because only C-merlin-342–595 could not be dephosphorylated on serum withdrawal (Fig. 3c, and Supplementary Fig. S3).

Dephosphorylation of merlin is important for its tumour suppressor function<sup>5,6</sup>. Several human missense mutations associated with neurofibromatosis type 2 have been found in the MYPT-1-interacting region of merlin<sup>14–17</sup>. To test whether these mutations affect the interaction with the phosphatase, we generated GST–C-merlin constructs carrying one of the naturally occurring mutations L316F, Q324L and L339F. One mutant, L339F, found in a patient with sporadic meningioma<sup>17</sup>, was indeed defective in MYPT-1 binding (Fig. 3d).



**Figure 2 | MYPT-1 is the targeting subunit of merlin phosphatase.** **a**, Co-immunoprecipitation (IP) of MYPT-1 and PP1 $\delta$  with merlin antibody (A-19). Immunoblots for MYPT-1 and PP1 $\delta$ . **b**, Pull-down of purified recombinant C-merlin with GST–C-MYPT-1. **c**, Co-immunoprecipitations of wild-type and mutant merlin (S518A or S518D) with MYPT-1 from cell lysates. **d**, *In vitro* dephosphorylation of merlin by the MYPT-1–PP1 $\delta$  complex immunoprecipitated with anti-MYPT-1. **e**, Trypsinolysis of MYPT-1–PP1 $\delta$  strongly reduces the catalytic function (phospho-merlin added after trypsin inhibitor for the periods indicated).



**Figure 3 | Interaction domains of merlin and MYPT-1.** **a**, GST–merlin constructs. **b**, GST–C-merlin deletion mutant pull-downs from cell lysates of MYPT-1 detected by immunoblot. **c**, C-merlin lacking the interaction region (amino-acid positions 312–341) is not dephosphorylated *in vivo* in response to serum withdrawal (immunoblot with phospho-merlin-specific antibodies). **d**, GST pull-downs of missense mutations in the MYPT-1 interacting region from a patient with neurofibromatosis type 2. **e**, The MYPT-1-binding-deficient L339F mutant does not dephosphorylate merlin *in vivo*. **f**, Scheme of MYPT-1 leucine zipper (LZ) wild type (WT) and mutant (MT). **g**, The MYPT-1 leucine zipper mutation shown in **f** destroys the interaction with merlin (Protein stain; LZ and WT migrate differently.)



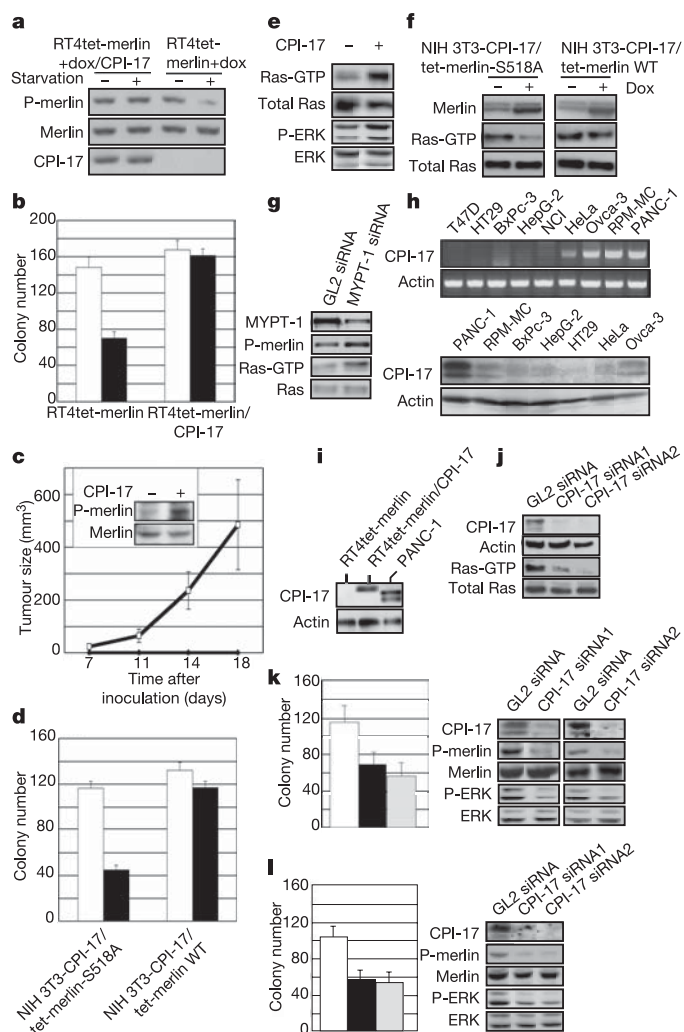
The disruption of the interaction with the activating phosphatase and loss of the subsequent dephosphorylation of the L339F mutant protein (Fig. 3e) could well explain its tumour-promoting effect.

MYPT-1 has a leucine zipper domain at its C terminus (Fig. 3f), which would be an ideal partner for the coiled-coil region of merlin. Mutation of the leucine residues to alanine (MYPT-1 L/A; 4) completely abolished MYPT-1 binding to other known substrates<sup>18–20</sup>. The GST-C-MYPT-1 L/A mutant, in contrast with the wild type, could not bind merlin (Fig. 3g). MYPT-1-associated PP1 $\delta$  therefore carries at least two binding sites for substrate, the N terminus for myosin<sup>9,10</sup> and the C-terminal leucine zipper domain for merlin.

Phosphatase activities need to be regulated; for example, MYPT-1–PP1 $\delta$  is inactivated by phosphorylation. Accordingly, MYPT-1 is dephosphorylated in the same way as merlin after serum withdrawal (Supplementary Fig. S4). In addition, MYPT-1–PP1 $\delta$  is counter-regulated by several cellular inhibitors<sup>10,21</sup>, namely phosphatase holoenzyme inhibitor (PH-1)<sup>22</sup>, kinase-enhanced phosphatase inhibitor (KEP-1)<sup>23</sup> and the 17-kDa protein kinase C potentiated inhibitor (CPI-17), the most specific and potent MYPT-1–PP1 $\delta$  inhibitor<sup>10,24</sup>. Overexpression of CPI-17 in RT4tet-merlin cells prevented merlin S518 dephosphorylation induced by serum withdrawal (Fig. 4a). Thus, inhibition of MYPT-1–PP1 $\delta$  by CPI-17 and thus of merlin activation inhibited the suppression of anchorage-independent growth (Fig. 4b).

To study the CPI-17-mediated inhibition of merlin dephosphorylation in non-tumorigenic cells, we chose NIH 3T3 cells, which contain no detectable CPI-17. CPI-17 expression in NIH 3T3 cells increased the level of merlin phosphorylation (Fig. 4c, inset) and, as expected, that of moesin, a known substrate of MYPT-1–PP1 $\delta$ <sup>8</sup> (data not shown). In addition, CPI-17 expression and subsequent inhibition of MYPT-1–PP1 $\delta$ , already known as an actomyosin contractility regulator, promoted cell shape changes and increased motility (data not shown). The expression of CPI-17 in NIH 3T3 cells caused all aspects of tumorigenic transformation: focus formation in the confluent monolayer culture and colony growth in soft agar (Supplementary Fig. S5a), and the formation of subcutaneous tumours when injected into nude mice (Fig. 4c, and Supplementary Fig. S5b). Because MYPT-1–PP1 $\delta$  acts on several other substrates besides merlin that could be responsible for tumour suppression, we attempted to rescue cells from CPI-17-induced transformation by expressing constitutively active merlin (S518A) under a doxycycline-inducible promoter. Merlin S518A, but not wild-type merlin, inhibited the growth of agar colonies despite overexpression of CPI-17 (Fig. 4d), indicating that MYPT-1–PP1 $\delta$  suppresses the tumorigenic phenotype by activating merlin. How the inhibition of MYPT-1–PP1 $\delta$  creates tumorigenic properties could be explained by knowing that merlin interferes with the activation of Ras (H.M., T.S. and P.H., unpublished observations); in agreement, inhibition of MAP kinase/ERK kinase blocked CPI-17-dependent transformation: Supplementary Fig. S5c). Thus, expression of CPI-17 in NIH 3T3 cells released Ras from inhibition by merlin and caused elevated levels of active Ras-GTP and phospho-ERK (Fig. 4e). This was prevented by introduction of the constitutively active S518A merlin mutant (Fig. 4f). The causal chain CPI-17–MYPT-1–merlin–Ras was further documented by down-regulating MYPT-1 by short interfering RNA (siRNA), which also blocked merlin dephosphorylation and increased Ras activation (shown for HeLa cells in Fig. 4g; not shown for RPM-MC cells).

From gene disruption in the mouse<sup>2–3</sup> we know that merlin is an essential gene. In addition, tumours in humans often carry merlin mutations<sup>25–27</sup>. On the basis of our data, elevated CPI-17 expression in tumorigenesis would be an alternative to loss of merlin function due to mutation. We found several human tumour cell lines with elevated CPI-17 mRNA (Fig. 4h, upper panel) and protein expression (Fig. 4h, lower panel). The abundance of CPI-17 protein in some tumour cell lines was almost as great as the ectopic expression level in both RT4tet-merlin/CPI-17 and NIH 3T3–CPI-17 cells (Fig. 4i, and not shown). To show that CPI-17 is involved in human tumorigenesis, we



**Figure 4 | The MYPT-1–PP1 $\delta$ -specific inhibitor CPI-17 inhibits merlin activation, promotes Ras activation and induces cellular transformation.**

**a, b**, CPI-17 expression prevents merlin dephosphorylation on serum withdrawal (**a**; immunoblots for phospho-S518 merlin, CPI-17 and bulk merlin) and merlin-dependent reduction of colony growth on soft agar (**b**; means  $\pm$  s.d. of three independent experiments;  $P < 0.01$ , Student's *t*-test). Open bars in **b**, no doxycycline; filled bars, with doxycycline. **c**, CPI-17 (open symbols) increases merlin phosphorylation (the inset shows an immunoblot for phospho-S518 merlin in NIH 3T3 cells) and transforms NIH 3T3 cells as shown by tumour growth in nude mice. Filled symbols, NIH 3T3 cells. Results are means  $\pm$  s.e.m. of tumour volumes from four animals. **d, f**, CPI-17-promoted agar colony growth (**d**; means  $\pm$  s.d. of at least three independent experiments;  $P < 0.01$ ) and Ras activation (**f**; method as in **e**) is inhibited by doxycycline-induced mutant merlin S518A (which does not require dephosphorylation), but not the wild-type (WT) merlin. Open bars in **d**, no doxycycline; filled bars, with doxycycline. **e**, CPI-17 increases Ras activity in NIH 3T3 cells. Ras-GTP was precipitated with Raf-1RBD; immunoblots were performed with the antibodies indicated. **g**, MYPT-1 downregulation by siRNA increases merlin phosphorylation and promotes Ras activation in HeLa cells. **h**, Expression of CPI-17 in different human tumour cell lines (upper panel, RT-PCR; lower panel, immunoblot). **i**, Comparison of ectopic Flag-tagged CPI-17 expression in RT4tet-merlin/CPI-17 cells with that in PANC-1 cells. **j**, siRNA downregulation of CPI-17 in RPM-MC cells decreases Ras activation. **k, l**, CPI-17 knockdown by two different siRNAs (black bars, siRNA1; grey bars, siRNA2) inhibits colony growth on agar and ERK activation in PANC-1 cells (**k**) and RPM-MC cells (**l**) compared with control luciferase (GL2; white bars). Quantitative results are means  $\pm$  s.d. of at least three independent experiments;  $P < 0.01$ . Lysates were immunoblotted as indicated.

downmodulated CPI-17 by siRNA in two human tumour cell lines, the pancreatic cell line PANC-1 and the melanoma cell line RPM-MC, and tested agar colony growth as a measure of the transformed phenotype. The colony number was indeed significantly decreased by two different CPI-17 specific siRNAs (Fig. 4k, l). As expected, the CPI-17 siRNA decreased the level of merlin phosphorylation and consequently Ras and ERK activity in both tumour cell lines (Fig. 4j–l).

Our observations show that MYPT-1–PP1 $\delta$  is important for cellular growth control in that it activates the tumour suppressor protein merlin, which subsequently targets the Ras pathway. Transformation by the oncoprotein CPI-17 demonstrates that the elimination of PP1 activity shown here and that of PP2A<sup>28</sup> both contribute to tumorigenesis. Although merlin inactivation is a major determinant of transformation by CPI-17, we also find elevated phosphorylation of ezrin, radixin and moesin (ERM; data not shown). Active phosphorylated ERM proteins are probably essential for Ras activation and proliferation control (H.M., T.S. and P.H., unpublished observations). Because merlin antagonizes ERM function<sup>5</sup>, CPI-17 probably induces a tumour-promoting activity of ERMs and loss of the counteracting role of merlin. Taken together, our data form the basis for proposing a new pathway leading to transformation that includes a phosphatase and an inhibitor involved in actomyosin contractility.

## METHODS

**Enzymes and other materials.** These are described in Supplementary Information.

**Antibodies and plasmids.** Flag antibody (M2) was obtained from Sigma; phospho-ERK (Thr202/Tyr204) antibody from Cell Signalling; CPI-17 antibodies from Santa Cruz and Upstate; MYPT-1 antibodies from Santa Cruz for immunoprecipitations and Becton Dickinson for immunoblotting; phospho-S518–merlin antibody from Rockland; phospho-MYPT-1 (Thr 696), Erk1 (K23), merlin (C-18 and A-19), actin (I-19), pan PP1 (E-9), pan PP2A (C-20) and PP1 $\delta$  (also called  $\beta$ N-9) antibodies from Santa Cruz; and Ras antibody (Ras10) from Upstate. GST–C-MYPT-1-expressing vector (wild-type and mutated type) was a gift from H. K. Surks. CPI-17-expressing vector was provided by A. Aitken. *Eco*RI/ fragments encoding C-merlin cDNA (residues 299–595, 312–595, 342–595 and 507–595) were amplified from pUHD10-3-NF-2 and subcloned into pGEX 4T.1 to produce GST–C-merlin.

**Cell culture.** Cells were maintained in DMEM medium supplemented with 10% FCS. Human tumour cell lines are described in Supplementary Information.

**Transfections.** Transfections were performed with Eugene 6 (Roche), in accordance with the manufacturer's instructions. The generation of stable clones is described in Supplementary Information.

**siRNA.** Sequences of siRNA are described in Supplementary Information. Gene silencing was achieved by transfection with 20 nM siRNA by using Oligofectamine (Invitrogen) in accordance with the manufacturer's instructions.

**SDS–PAGE, western blotting and Coomassie brilliant blue staining.** Procedures are described in Supplementary Information.

**RT–PCR.** The procedure is described in Supplementary Information.

**Immunoprecipitation and GST pull-down assay.** The procedure is described in Supplementary Information.

**In vitro dephosphorylation assay.** The procedure is described in Supplementary Information.

**Trypsinolysis of MYPT-1.** Immunoprecipitated MYPT-1–PP1 $\delta$  complex was mixed with an equal volume of either 1 mg ml<sup>−1</sup> trypsin (trypsinolysis) or 20 mg ml<sup>−1</sup> soybean trypsin inhibitor (control trypsinolysis) and incubated for 5 min at 30 °C. Incubations were terminated with an equal volume of 20 mg ml<sup>−1</sup> soybean trypsin inhibitor or 1 mg ml<sup>−1</sup> trypsin, respectively. The merlin substrate was incubated at 30 °C for 5–20 min. Phosphorylated merlin was prepared as described in Supplementary Information.

**Soft agar assay.** The procedure is described in Supplementary Information.

**Tumour growth in vivo.** The procedure is described in Supplementary Information.

Received 21 October 2005; accepted 2 May 2006.

- McClatchey, A. I. Merlin and ERM proteins: unappreciated roles in cancer development? *Nature Rev. Cancer* **3**, 877–883 (2003).
- McClatchey, A. I. *et al.* Mice heterozygous for a mutation at the *Nf2* tumor suppressor locus develop a range of highly metastatic tumors. *Genes Dev.* **12**, 1121–1133 (1998).
- McClatchey, A. I., Saotome, I., Ramesh, V., Gusella, J. F. & Jacks, T. The *Nf2*

tumor suppressor gene product is essential for extraembryonic development immediately prior to gastrulation. *Genes Dev.* **11**, 1253–1265 (1997).

- Xiao, G. H., Chernoff, J. & Testa, J. R. NF2: the wizardry of merlin. *Genes Chromosomes. Cancer* **38**, 389–399 (2003).
- Morrison, H. *et al.* The NF2 tumor suppressor gene product, merlin, mediates contact inhibition of growth through interactions with CD44. *Genes Dev.* **15**, 968–980 (2001).
- Shaw, R. J. *et al.* The *Nf2* tumor suppressor, merlin, functions in Rac-dependent signalling. *Dev. Cell* **1**, 63–72 (2001).
- Downay, A. B. & Forsyth, C. J. Okadaic acid: the archetypal serine/threonine protein phosphatase inhibitor. *Curr. Med. Chem.* **9**, 1939–1980 (2002).
- Fukata, Y. *et al.* Association of the myosin-binding subunit of myosin phosphatase and moesin: dual regulation of moesin phosphorylation by Rho-associated kinase and myosin phosphatase. *J. Cell Biol.* **141**, 409–418 (1998).
- Hartshorne, D. J. & Hirano, K. Interactions of protein phosphatase type 1, with a focus on myosin phosphatase. *Mol. Cell. Biochem.* **190**, 79–84 (1999).
- Ito, M., Nakano, T., Erdodi, F. & Hartshorne, D. J. Myosin phosphatase: structure, regulation and function. *Mol. Cell. Biochem.* **259**, 197–209 (2004).
- Beullens, M., Van Eynde, A., Stalmans, W. & Bollen, M. The isolation of novel inhibitory polypeptides of protein phosphatase 1 from bovine thymus nuclei. *J. Biol. Chem.* **267**, 16538–16544 (1992).
- Colbran, R. J. *et al.* Association of brain protein phosphatase 1 with cytoskeletal targeting/regulatory subunits. *J. Neurochem.* **69**, 920–929 (1997).
- Rouleau, G. A. *et al.* Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. *Nature* **363**, 515–521 (1993).
- Shimizu, T. *et al.* Structural basis for neurofibromatosis type 2. Crystal structure of the merlin FERM domain. *J. Biol. Chem.* **277**, 10332–10336 (2002).
- Baser, M. E. *et al.* Genotype–phenotype correlations for nervous system tumors in neurofibromatosis 2: a population-based study. *Am. J. Hum. Genet.* **75**, 231–239 (2004).
- Baser, M. E. *et al.* Presymptomatic diagnosis of neurofibromatosis 2 using linked genetic markers, neuroimaging, and ocular examinations. *Neurology* **47**, 1269–1277 (1996).
- De Vitis, L. R. *et al.* Screening for mutations in the neurofibromatosis type 2 (NF2) gene in sporadic meningiomas. *Hum. Genet.* **97**, 632–637 (1996).
- Surks, H. K., Richards, C. T. & Mendelsohn, M. E. Myosin phosphatase–Rho interacting protein. A new member of the myosin phosphatase complex that directly binds RhoA. *J. Biol. Chem.* **278**, 51484–51493 (2003).
- Surks, H. K. & Mendelsohn, M. E. Dimerization of cGMP-dependent protein kinase 1 $\alpha$  and the myosin-binding subunit of myosin phosphatase: role of leucine zipper domains. *Cell. Signal.* **15**, 937–944 (2003).
- Surks, H. K. *et al.* Regulation of myosin phosphatase by a specific interaction with cGMP-dependent protein kinase 1 $\alpha$ . *Science* **286**, 1583–1587 (1999).
- Feng, J. *et al.* Inhibitory phosphorylation site for Rho-associated kinase on smooth muscle myosin phosphatase. *J. Biol. Chem.* **274**, 37385–37390 (1999).
- Eto, M., Karginov, A. & Brautigan, D. L. A novel phosphoprotein inhibitor of protein type-1 phosphatase holoenzymes. *Biochemistry* **38**, 16952–16957 (1999).
- Liu, Q. R. *et al.* KEPI, a PKC-dependent protein phosphatase 1 inhibitor regulated by morphine. *J. Biol. Chem.* **277**, 13312–13320 (2002).
- Eto, M., Senba, S., Morita, F. & Yazawa, M. Molecular cloning of a novel phosphorylation-dependent inhibitory protein of protein phosphatase-1 (CPI17) in smooth muscle: its specific localization in smooth muscle. *FEBS Lett.* **410**, 356–360 (1997).
- Evans, D. G., Sainio, M. & Baser, M. E. Neurofibromatosis type 2. *J. Med. Genet.* **37**, 897–904 (2000).
- Gutmann, D. H. New insights into the neurofibromatoses. *Curr. Opin. Neurol.* **7**, 166–171 (1994).
- Uppal, S. & Coatesworth, A. P. Neurofibromatosis type 2. *Int. J. Clin. Pract.* **57**, 698–703 (2003).
- Hahn, W. C. & Weinberg, R. A. Modelling the molecular circuitry of cancer. *Nature Rev. Cancer* **2**, 331–341 (2002).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank N. Howells for the animal experiments performed in the Institute of Toxicology and Genetics at the Forschungszentrum Karlsruhe, and F. Böhmer and C. Calkhoven for helpful discussions. This work was supported by the Deutsche Forschungsgemeinschaft (DFG), the Association for International Cancer Research and the Sonderforschungsbereich (SFB).

**Author Contributions** H.M. was the project leader. H.M. and H.J. were responsible for the experimental and project design. H.J. performed most the experiments. H.M. and T.S. performed some of the experiments. H.M. and P.H. wrote the manuscript and all authors discussed the results and contributed to the writing of the manuscript. H.M. and T.S. prepared the final version of the manuscript.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.M. ([helen@fli-leibniz.de](mailto:helen@fli-leibniz.de)).

## LETTERS

# Spatial control of actin organization at adherens junctions by the synaptotagmin-like protein Btsz

Fanny Pilot<sup>1</sup>, Jean-Marc Philippe<sup>1</sup>, Céline Lemmers<sup>1</sup> & Thomas Lecuit<sup>1</sup>

Epithelial tissues maintain a robust architecture during development. This fundamental property relies on intercellular adhesion through the formation of adherens junctions containing E-cadherin molecules<sup>1,2</sup>. Localization of E-cadherin is stabilized through a pathway involving the recruitment of actin filaments by E-cadherin<sup>3–6</sup>. Here we identify an additional pathway that organizes actin filaments in the apical junctional region (AJR) where adherens junctions form in embryonic epithelia. This pathway is controlled by Bitesize (Btsz), a synaptotagmin-like protein<sup>7,8</sup> that is recruited in the AJR independently of E-cadherin and is required for epithelial stability in *Drosophila* embryos. On loss of *btsz*, E-cadherin is recruited normally to the AJR, but is not stabilized properly and actin filaments fail to form a stable continuous network. In the absence of E-cadherin, actin filaments are stable for a longer time than they are in *btsz* mutants. We identify two polarized cues that localize Btsz: phosphatidylinositol (4,5)-bisphosphate, to which Btsz binds; and Par-3. We show that Btsz binds to the Ezrin–Radixin–Moesin protein Moesin, an F-actin-binding protein that is localized apically<sup>9</sup> and is recruited in the AJR in a *btsz*-dependent manner. Expression of a dominant-negative form of Ezrin that does not bind F-actin phenocopies the loss of *btsz*. Thus, our data indicate that, through their interaction, Btsz and Moesin may mediate the proper organization of actin in a local domain, which in turn stabilizes E-cadherin. These results provide a mechanism for the spatial order of actin organization underlying junction stabilization in primary embryonic epithelia.

Homotypic binding of the cell-adhesion molecule E-cadherin (E-cad) at the adherens junctions of epithelial cells organizes the formation of multiprotein complexes, composed in part of the  $\beta$ -catenin and  $\alpha$ -catenin proteins, and their dynamic interaction with actin filaments (F-actin)<sup>1,3,5,6</sup>. F-actin is required to stabilize E-cad– $\beta$ -catenin– $\alpha$ -catenin complexes<sup>4</sup>. Moreover, E-cad regulates its own stability through the organization of actin filaments through  $\alpha$ -catenin:  $\alpha$ -catenin binds Formin (also known as Diaphanous)<sup>10</sup> and suppresses branching by competing with Arp2/3 (ref. 5). When epithelia form through the mesenchymal–epithelial transition, the sites of initial cell contact serve as spatial landmarks for the recruitment of E-cad– $\beta$ -catenin– $\alpha$ -catenin complexes during the formation of adherens junctions<sup>11</sup>. In primary embryonic epithelia, however, adherens junctions do not form through specific cell contacts, and the spatial cues positioning the adherens junctions in the AJR are less characterized and may be different<sup>2,11,12</sup>. The identification of such spatial cues and the mechanisms whereby these cues organize structural, cytoskeletal components associated with the formation and/or stabilization of adherens junctions is an important challenge.

We have addressed this problem in the early *Drosophila* embryo. Formation, stabilization and remodelling of adherens junctions occur in a tightly and genetically controlled sequence involving *e-cad* (or *shotgun*)<sup>13,14</sup>,  $\beta$ -catenin (or *armadillo*)<sup>15,16</sup>, *par-6* (ref. 17),

*par-3* (or *bazooka*, *baz*)<sup>18</sup>, *aPKC*<sup>19</sup>, *crumbs*<sup>20</sup> and others. A microarray-based RNA interference (RNAi) screen of epithelial morphogenesis<sup>21</sup> identified *btsz*, a gene previously known to control growth in adult flies<sup>8</sup>, as a regulator of embryonic epithelial integrity. In embryos injected with double-stranded RNA (dsRNA) probes specific for *btsz* (hereafter called *btszRNAi* embryos), gastrulation was severely affected and the epithelium failed to extend properly (Supplementary Fig. S1 and Movies M1–M3). Defects were either strong or medium; that is, they were visible at the beginning of gastrulation (Supplementary Movie M2) or about 15 min later, respectively (Supplementary Movie M3). The defects were penetrant (80%) and dose dependent (Supplementary Fig. S2a). Four different, nonoverlapping probes produced these defects and embryos were not affected with control probes (Supplementary Fig. S2a).

Btsz is the only *Drosophila* member of the synaptotagmin-like protein (SLP) family<sup>7,8</sup> characterized by the presence of tandem carboxy-terminal C2 boxes. *btsz* encodes several isoforms<sup>8</sup> (Supplementary Fig. S2b). In early embryos, *btsz1* was not detected but *btsz2* and *btsz3* were expressed (on the basis of northern blots, Fig. S2c, and polymerase chain reaction with reverse transcription (RT–PCR); data not shown) together with *btsz0*, another isoform not previously reported<sup>8</sup> (Supplementary Fig. S2b, c). We found that at least one of these isoforms is maternally and zygotically provided (Supplementary Fig. S3). The most efficient dsRNA probes used recognized all three maternally and zygotically expressed isoforms (Supplementary Figs S2b, c, and S3). These isoforms were strongly reduced in *btszRNAi* embryos (Supplementary Fig. S4), suggesting that RNAi produces a severe *btsz* loss-of-function phenotype.

Two *btsz* alleles have been described<sup>8</sup>: *btszK13-4* introduces a deletion in the amino terminus of *btsz2* (residues 501–1,494), *btszJ5-2* corresponds to a frameshift mutation that introduces a stop codon at amino acid 390, which leads to a truncation in Btsz0 and Btsz2, and probably the absence of Btsz3. *btszK13-4* homozygous female escapers can be recovered and were crossed to heterozygous *btszK13-4* or *btszJ5-2* males. Although many embryos were not fertilized, those that were reached cellularization and showed epithelial defects during gastrulation: 26% ( $n = 23$ ) of *btszK13-4/btszK13-4* and 46% ( $n = 22$ ) of *btszK13-4/btszJ5-2* embryos (Supplementary Figs S1d–f, S2f, and Movies M4–M6). We found that *btszJ5-2* germline clones do not produce eggs (data not shown) and that *btszJ5-2* is lethal, in contrast to a previous report<sup>8</sup>. We examined trans-heterozygous embryos with a deficiency removing the *btsz* locus (*Df(3R)Exel6275*, called *Dfbtsz*): 12% ( $n = 18$ ) of embryos from crosses of *Dfbtsz/btszK13-4* females and wild-type males showed epithelial defects. This proportion went up to 39% when males were heterozygous *btszK13-4/+*. We conclude that *btsz* is zygotically and maternally required. Whereas RNAi targeted all three *btsz* isoforms, *btszK13-4* left intact a large fraction of Btsz2 and Btsz0, probably explaining the weaker penetrance of phenotypes in *btszK13-4* (26%) or

<sup>1</sup>Institut de Biologie du Développement de Marseille Luminy (IBDML) UMR 6216, CNRS-Université de la Méditerranée. Campus de Luminy, case 907, 13288 Marseille cedex 09, France.

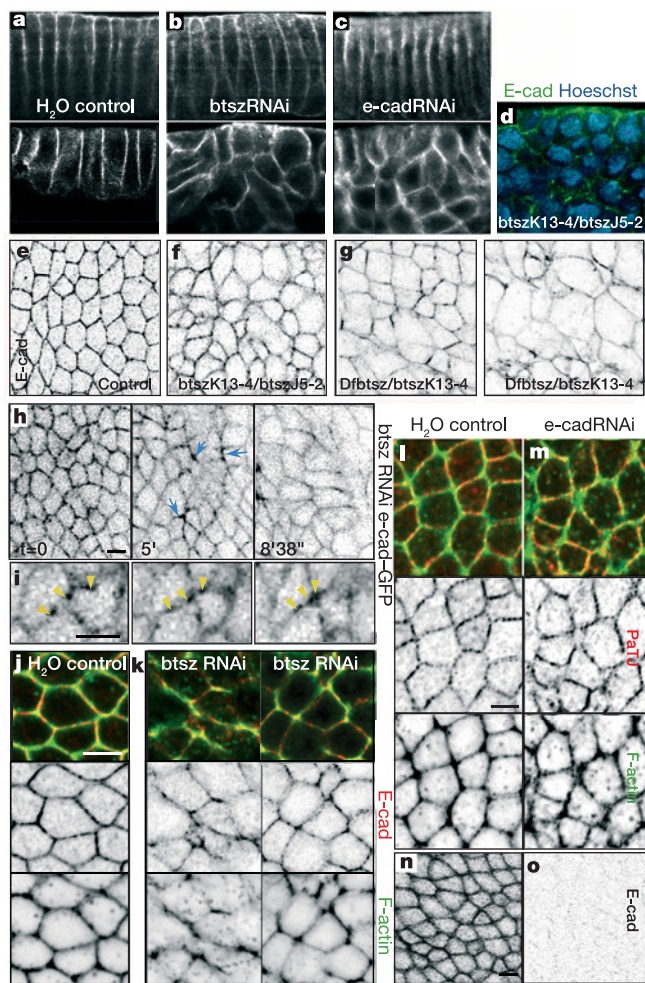


*btszK13-4/btszJ5-2* (46%) mutants, as compared with *btszRNAi* embryos (80%). Notably, despite its essential role in the formation of epithelia in early embryos, the recovery of adult escapers suggests that *btsz* may be dispensable in adult epithelia.

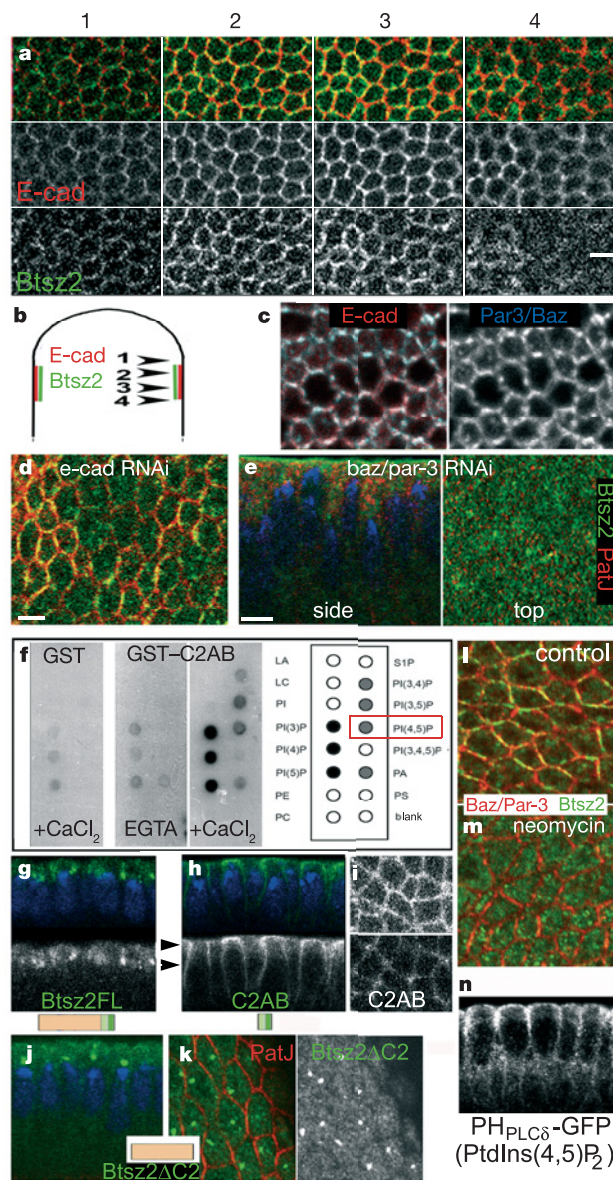
Overexpression of a *btsz2* isoform lacking the 3' untranslated region (UTR) rescued the phenotypes produced by an RNAi probe targeting the 3' UTR of all *btsz* isoforms (Supplementary Fig. S2d). Overexpression of *btsz2* more robustly rescued the *btszRNAi* phenotype than did *btsz3* overexpression (data not shown), suggesting that *btsz2* has a prominent role. The injection of morpholino antisense oligonucleotides (morpholinos) specific to each *btsz* isoform confirmed this (Supplementary Fig. S2e): a control morpholino showed no defect (<2%,  $n = 58$ ), a mix of *btsz0*, *btsz2* and *btsz3* morpholinos caused penetrant defects (92%,  $n = 103$ ), and a *btsz2*-specific morpholino alone caused defects in 73% of embryos ( $n = 58$ ). We

therefore focused the following experiments on *Btsz2*, a 286-kDa protein (2,645 residues).

The expression of a Glu-epitope-tagged variant of *Btsz2* (*Btsz2-Glu*<sup>8</sup>) was strongly reduced in *btszRNAi* embryos (data not shown). The epithelium failed to maintain its regular morphology in *btszRNAi* embryos (Supplementary Fig. S1a–c), *btsz* mutants (Supplementary Fig. S1e, f) and *btsz* morphants (data not shown). Although cellularization proceeded similarly in *btszRNAi* and control embryos, at the onset of gastrulation the epithelium collapsed and became multilayered in *btszRNAi* (Fig. 1b) and *btszK13-4/btszJ5-2*



**Figure 1 | Epithelial integrity, adherens junctions stability and actin organization in embryos deficient for *btsz* or *e-cad*.** **a–d**, Side views of the epithelium labelled with Neurotactin (**a–c**) or E-cad (**d**) are shown at the end of cellularization (top) and during gastrulation (bottom) in control (**a**), *btszRNAi* (**b**), *e-cadRNAi* (**c**) and *btszK13-4/btszJ5-2* (**d**) embryos. **e–g**, Top views of adherens junctions marked with an antibody against E-cad in control (**e**), *btszK13-4/btszJ5-2* (**f**) and *Dfbtsz/btszK13-4* (**g**, left and right) embryos. **h, i**, E-cadherin-GFP in a living *btszRNAi* embryo disappears progressively in the AJR (**h**), forming clusters (**h**, arrow; **i**, arrowheads). **j, k**, Adherens junctions labelled with E-cad (red) and phalloidin (green). The actin belt is regular in controls (**j**), and fragmented in *btszRNAi* embryos (**k**, left and right). **l–o**, Adherens junctions marked with PatJ (**l, m**, red) and phalloidin (**l, m**, green) or E-cad (**n, o**, black). RNAi directed against *e-cad* inhibits the recruitment of E-cad at adherens junctions (**n, o**) and does not markedly disrupt the actin belt (**l, m**). Scale bar, 5  $\mu$ m.



**Figure 2 | *Btsz2* localization in the AJR requires *Par-3* and *PtdIns(4,5)P<sub>2</sub>*.** **a, b**, Four sections spanning the AJR labelled with E-cad (red) and *Btsz2-Glu* (green). **c**, Colocalization of E-cad (red) and *Par-3/Baz* (blue) in the AJR. **d**, *Btsz2-Glu* (green) and *PatJ* (red) localization in the AJR of *e-cadRNAi* (**d**) and *par-3RNAi* embryos (**e**). **f**, Phospholipid-binding assays with GST and GST-C2AB with (+CaCl<sub>2</sub>) or without (EGTA) Ca<sup>2+</sup>. LA, lysophosphatidic acid; PA, phosphatidic acid; PS, PC, PE, PI, phosphatidyl-serine, -choline, -ethanolamine and -inositol; S1P, sphingosine 1-phosphate. **g–k**, Localization of *Btsz2-Glu* full-length (*Btsz2FL*, **g**) or C2AB-HA (**h, i**) and *Btsz2-ΔC2-HA* (**j, k**) in embryos marked with Hoechst (**g–j**, blue) or *PatJ* (**k**, red) in side (**g, h, j**) or top (**i, k**) views. **l, m**, *Btsz2-Glu* (green) and *Par-3* (red) localization in neomycin injected (**m**) and control embryos (**l**). **n**, Localization of PH<sub>PLC6</sub>-GFP in early embryos. Scale bar: 5 microns.

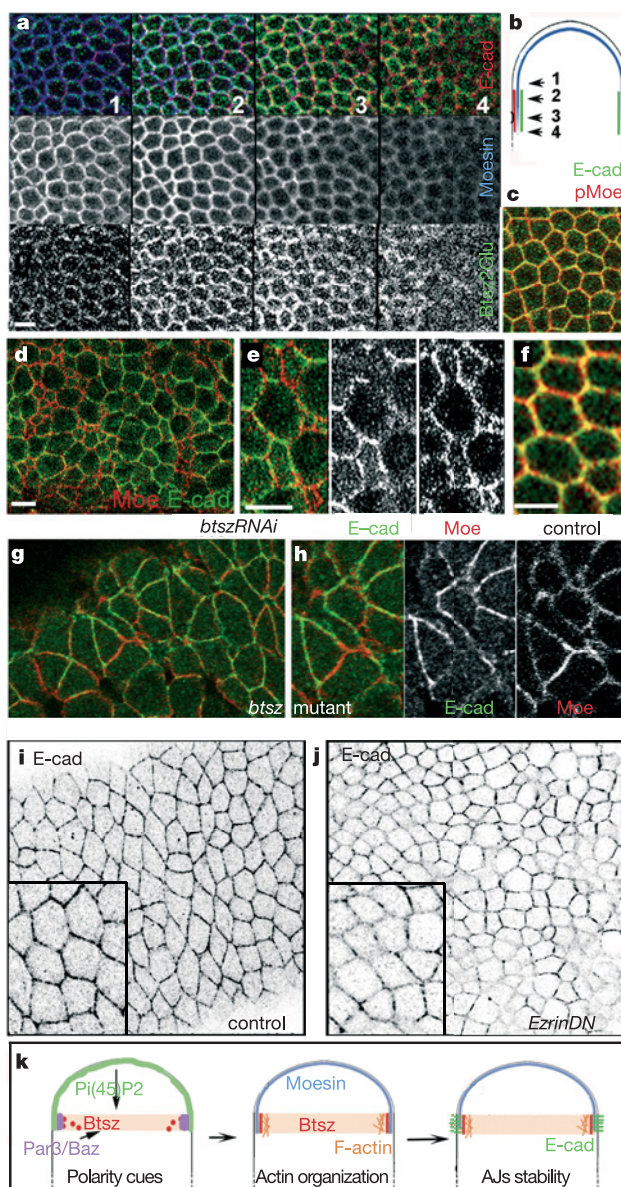


mutant (Fig. 1d) embryos, as compared with controls (Fig. 1a). A similar phenotype was observed in *e-cadRNAi* embryos (Fig. 1c). Thus, *btsz* controls the stable architecture of primary embryonic epithelia.

These data suggested that *btsz* might regulate the formation of adherens junctions. In contrast to the wild type, in which E-cad was uniformly present at the adherens junctions (Fig. 1e), E-cad expression was heterogeneous and the adherens junctions appeared severely fragmented in *btsz* mutants (Fig. 1f, g) and *btszRNAi* embryos (Fig. 1h–k). Time-lapse recordings of E-cad fused to green fluorescent protein (GFP) showed that adherens junctions formed properly in the AJR of *btszRNAi* embryos (Fig. 1h) but that, subsequently, E-cad–GFP expression disappeared (Fig. 1h and Supplementary Movie M7), suggesting a defect in the stabilization but not targeting of E-cad. E-cad–GFP, or endogenous E-cad, disappeared in small patches at cell contacts (Fig. 1i–k), pointing to defects in actin organization. Indeed, the actin belt in the AJR was fragmented in *btszRNAi* embryos (Fig. 1j, k). We tested whether actin organization or the E-cad– $\beta$ -catenin– $\alpha$ -catenin complexes was the primary cause of the disassembly of adherens junctions in *btszRNAi* embryos. In *e-cadRNAi* embryos, in which E-cad was undetectable in the nascent AJR (Fig. 1n, o), the actin belt was not considerably affected during early gastrulation (Fig. 1l, m) and clearly less affected than in *btszRNAi* embryos at the same stage (Fig. 1k). Subsequently, however, F-actin was disorganized in *e-cadRNAi* embryos (data not shown). This suggests that Btsz is part of an E-cad-independent pathway controlling actin organization in the AJR and consequently junction stability.

We next examined Btsz2 localization. Btsz2–Glu is a functional protein that rescues the *btszRNAi* phenotype (Supplementary Fig. S2d). Btsz2–Glu was previously reported to localize apically in follicular epithelial cells<sup>8</sup>. In early embryos, Btsz2–Glu was detected at

the AJR together with E-cad from the end of cellularization until about 30 min into gastrulation (Fig. 2a, b). Subsequently, Btsz2–Glu was found in a subapical compartment (data not shown). At these early stages, E-cad colocalized with Par-3 (also known as Baz) (Fig. 2c). We thus addressed the possible role of E-cad and Par-3/Baz in Btsz2 localization in the AJR. In *e-cadRNAi* embryos, the recruitment of E-cad in the AJR was blocked (Fig. 1o) and Btsz2 was normal (Fig. 2d); by contrast, in *par-3/bazRNAi* embryos Btsz2 was largely cytoplasmic (Fig. 2e), like PatJ, another marker of AJR at this stage. Btsz2 is thus a target of the early polarity marker Par-3/Baz, which is required for E-cad localization in the AJR<sup>12</sup>.



**Figure 3 | Btsz interacts with Moesin.** **a, b**, Structure of Moesin (**a**) and Btsz2 (**b**), indicating the domains used for the yeast-two-hybrid assay. ABD, actin-binding domain of Moesin; MBD, Moesin-binding-domain of Btsz2. MBDC contains the MBD and a coiled coil (see Methods). **c**, Yeast two-hybrid interactions. The seven combinations of interactions are shown on the left. DBD, DNA-binding domain; AD, activation domain. Yeast were streaked on media selecting for expression of the *HIS3* reporter gene. Quantification of liquid  $\beta$ -galactosidase assays (data are the mean  $\pm$  s.d. in arbitrary units). **d**, Immunoblots of GST pull-down assays using *Sophila* S2 cell lysates or embryonic extracts. Btsz2MBDC–HA was pulled down on GST–MoeFFF beads (left). Moesin, extracted from S2 cells or embryos, was pulled down on GST–Btsz2MBDC beads (right).

**Figure 4 | Btsz and Moesin cooperate to stabilize E-cad.** **a, b**, Four sections spanning the AJR, showing Moesin (blue), Btsz2–Glu (green) and E-cad (red). **c**, Colocalization of phosphorylated Moesin (red) and E-cad in the AJR. **d–h**, Localization of Moesin (red) and E-cad (green) in the AJR in control (**f**), *btszRNAi* (**d, e**) and *btsz* mutant (**g, h**) embryos. Scale bar, 5  $\mu$ m. **i, j**, Collapse of adherens junctions labelled with E-cad in embryos expressing EzrinDN (**j**) compared with control (**i**). Insets show close-up views. **k**, Model of the steps in the organization of adherens junctions: upstream polarity cues (Par-3 and PtdIns(4,5)P<sub>2</sub>) define a domain in the AJR where Btsz localizes (red). Btsz and Moesin interaction organizes actin filaments, which in turn stabilize E-cad.

We then tested the role of the two C2 boxes (C2AB) in the localization of Btsz2 (Fig. 2f–n). Purified glutathione S-transferase (GST)-tagged C2AB (unlike GST alone) bound to phosphatidylinositol mono- and bisphosphate species in a  $\text{Ca}^{2+}$ -dependent fashion *in vitro* (Fig. 2f). We assessed the *in vivo* relevance of this binding. A tagged form of Btsz2 lacking the C2 boxes (Btsz2- $\Delta\text{C2-HA}$ ) and expressed in gastrulating embryos was cytoplasmic and failed to localize at the AJR (Fig. 2j, k; compare with Fig. 2a, g). Conversely, an epitope-tagged form of C2AB (C2AB-HA) localized at the plasma membrane (Fig. 2h) in gastrulating embryos. Notably, C2AB was polarized and concentrated in the apical surface and in the AJR. Of all the phosphoinositides that C2AB binds *in vitro*, phosphatidylinositol (4,5)-bisphosphate ( $\text{PtdIns}(4,5)\text{P}_2$ ) is the most abundant at the plasma membrane, suggesting that  $\text{PtdIns}(4,5)\text{P}_2$  could be required for Btsz2 localization in the AJR. Injection of cellularizing embryos with neomycin, a compound that binds and inhibits  $\text{PtdIns}(4,5)\text{P}_2$  (ref. 22), resulted in epithelial defects similar to *btsz*, *par-3* or *e-cadRNAi* (data not shown), and inhibited the recruitment of Btsz2 at the plasma membrane and the AJR (Fig. 2l, m). A fusion between GFP and the pleckstrin homology (PH) domain of phospholipase C $\delta$  (PLC $\delta$ ), which specifically binds  $\text{PtdIns}(4,5)\text{P}_2$  (ref. 23), localized apically and in the AJR (Fig. 2n), similar to the Btsz C2 boxes (Fig. 2h). We conclude that  $\text{PtdIns}(4,5)\text{P}_2$  is a polarized spatial cue required for localization of Btsz in the AJR, and hence for adherens junction stability, together with Par-3/Baz.

How does localized Btsz organize F-actin in the AJR? A large-scale two-hybrid screen<sup>24</sup> identified an interaction between Btsz and Moesin, the only *Drosophila* member of the Ezrin–Radixin–Moesin (ERM) family of F-actin binding proteins<sup>9,25</sup> that has been implicated in various aspects of epithelial polarity. We confirmed this interaction and showed that Btsz bound to the third F3 subdomain of Moesin (Fig. 3a, c). We narrowed down a minimal region in Btsz that binds Moesin (Fig. 3b, c, Moesin-binding domain). This interaction occurred in GST pull-down assays of *Drosophila* S2 cell lysates and embryonic extracts (Fig. 3d). We next assessed the functional relevance of this interaction. Moesin and the phosphorylated active form of Moesin<sup>26</sup>, which binds F-actin, localized in early embryos in the apical surface and in the AJR, together with Btsz2 and E-cad (Fig. 4a–c), suggesting that the interaction between Btsz and Moesin may spatially define a domain of actin organization in the AJR required to stabilize E-cad. In agreement with this, in *btszRNAi* (Fig. 4d, e) and *btsz* mutant (Fig. 4g, h) embryos, Moesin localization was diminished in the AJR as compared with controls (Fig. 4f), and E-cad and Moesin segregated in distinct domains as E-cad progressively disappeared.

We tested whether Moesin is required for epithelial stability in early embryos. Moesin has a major maternal contribution and is a very stable protein. Moreover, females whose germline is mutant for *moesin* do not lay eggs<sup>27</sup>. Thus, we did not identify a phenotype using either various *moesin* mutant alleles or RNAi. We therefore expressed in early embryos a dominant-negative construct of Ezrin, a mammalian Moesin orthologue that lacks the C-terminal actin-binding domain and acts as a dominant-negative in *Drosophila* (EzrinDN, containing residues 1–280)<sup>28</sup>. We found that embryos expressing EzrinDN during gastrulation showed epithelial defects (41% of embryos,  $n = 115$ ; Fig. 4j) similar to *btsz* mutants and unlike control embryos (2% of embryos,  $n = 59$ ; Fig. 4i). In particular, cellularization was normal, adherens junctions formed properly, but E-cad was no longer present homogeneously around the AJR (Fig. 4i, j).

These results shed light on the mechanisms underlying the spatial control of actin filament and the stability of the adherens junctions in the *Drosophila* primary embryonic epithelium (Fig. 4k). In Btsz, we have identified an E-cad independent pathway that participates in F-actin organization in the AJR, together with Moesin. Btsz and Moesin bind and colocalize in the AJR in a *btsz*-dependent fashion, and expression of a mutant form of Ezrin that does not bind F-actin

disrupts adherens junctions stability similar to loss of *btsz*. Notably, this work identifies upstream polarity cues (Par-3/Baz and  $\text{PtdIns}(4,5)\text{P}_2$ ) that control the spatial order of actin organization at the AJR through the localization of Btsz. The fact that  $\text{PtdIns}(4,5)\text{P}_2$  acts as a key regulator of epithelial polarity in the early embryo raises the issue of how  $\text{PtdIns}(4,5)\text{P}_2$  metabolism is spatially regulated in epithelial cells. The observation that Par-3 binds PTEN<sup>29,30</sup>, which converts  $\text{PtdIns}(3,4,5)\text{P}_3$  into  $\text{PtdIns}(4,5)\text{P}_2$ , suggests that Par-3/Baz may be part of this process. We have thus identified a key intermediate between polarity cues and structural elements of adherens junctions important for embryonic epithelial stability. Five SLPs and two SLP-related (Slac2) proteins are close orthologues of Btsz in mammals<sup>7</sup>. It would be worth investigating their potentially conserved role in the dynamic organization of actin at adherens junctions in embryonic epithelia.

## METHODS

Detailed information on Methods is described in the Supplementary Information.

**Fly stocks.** Wild-type embryos were from *yw* and *OreR* stocks. We used the strain *ubi-E-cad-GFP (II)* (from H. Oda). Overexpression experiments were done by crossing females bearing the *67;15* driver (*mat  $\alpha\text{Tub-gal4 VP16}$  (67c); mat  $\alpha\text{Tub-gal4 VP16}$  (15)*) with males carrying *UAS-btsz2*, *UAS-btsz2Glu* (also called *btsz2Glu $\Delta$ 3' UTR* in Supplementary Fig. S2d or *btsz2-poly* in ref. 8), *UAS-btsz3* (from J. Serano), *UAS-Ezrin1–280* (gift from E. Suzuki), *UAS-C2AB-HA*, *UAS-btsz2- $\Delta\text{C2HA}$*  (this work), and *UAS-PH $\text{PLC}\delta$ -GFP* (from A. Zehhof). *w; btsz<sup>K13–4</sup>/TM3Sb KrGal4-UASGFP* and *w; FRT82B btsz<sup>15–2</sup>/TM3Sb KrGal4-UASGFP. Df(3R)Exel6275/TM3Sb* (Bloomington 7742) flies were used to delete the *btsz* locus (88D4–5; confirmed by genomic PCR, data not shown).

**Molecular biology.** Information on DNA constructs, RT-PCR and northern blots, yeast two-hybrid assays, GST fusion proteins, GST pull-down assays, lipid-binding assays and *Drosophila* S2 cells is given in the Supplementary Information.

**RNAi experiments.** RNAi was done as described<sup>21</sup>. All details on the sequence of probes are given in the Supplementary Information. Embryos were obtained from 30-min egg collections at 25 °C for wild-type stocks (or expressing E-cad-GFP) and from 1-h egg collections at 18 °C for embryos laid by *67;15* females crossed to males carrying *UAS* transgenes. Except for the experiments shown in Supplementary Fig. S1, all *btszRNAi* assays used probe 1.

**Morpholino injections.** Translation-blocking morpholino antisense oligonucleotides, synthesized by Gene Tools were used to knock-down the expression of the different *btsz* isoforms. Further information is given in the Supplementary Information. Early embryos were injected and recorded as described for RNAi. The control was injected at 2 mM yielding a 20–40  $\mu\text{M}$  final morpholino concentration and *btsz* specific morpholinos were injected at 0.33 mM each.

Embryos were fixed using standard protocols. The list of antibodies used, their origin and dilution conditions are given in Supplementary Information.

Received 28 February; accepted 16 May 2006.

Published online 9 July 2006.

- Gumbiner, B. M. Regulation of cadherin adhesive activity. *J. Cell Biol.* **148**, 399–404 (2000).
- Knust, E. & Bossinger, O. Composition and formation of intercellular junctions in epithelial cells. *Science* **298**, 1955–1959 (2002).
- Jamora, C. & Fuchs, E. Intercellular adhesion, signalling and the cytoskeleton. *Nature Cell Biol.* **4**, E101–E108 (2002).
- Bershadsky, A. Magic touch: how does cell–cell adhesion trigger actin assembly? *Trends Cell Biol.* **14**, 589–593 (2004).
- Drees, F., Pokutta, S., Yamada, S., Nelson, W. J. & Weis, W. I.  $\alpha$ -Catenin is a molecular switch that binds E-cadherin- $\beta$ -catenin and regulates actin-filament assembly. *Cell* **123**, 903–915 (2005).
- Yamada, S., Pokutta, S., Drees, F., Weis, W. I. & Nelson, W. J. Deconstructing the cadherin–catenin–actin complex. *Cell* **123**, 889–901 (2005).
- Fukuda, M., Saegusa, C. & Mikoshiba, K. Novel splicing isoforms of synaptotagmin-like proteins 2 and 3: identification of the SLP homology domain. *Biochem. Biophys. Res. Commun.* **283**, 513–519 (2001).
- Serano, J. & Rubin, G. M. The *Drosophila* synaptotagmin-like protein bitesize is required for growth and has mRNA localization sequences within its open reading frame. *Proc. Natl Acad. Sci. USA* **100**, 13368–13373 (2003).
- Bretscher, A., Edwards, K. & Fehon, R. G. ERM proteins and merlin: integrators at the cell cortex. *Nature Rev. Mol. Cell Biol.* **3**, 586–599 (2002).
- Kobiela, A., Pasolli, H. A. & Fuchs, E. Mammalian formin-1 participates in adherens junctions and polymerization of linear actin cables. *Nature Cell Biol.* **6**, 21–30 (2004).



11. Nelson, W. J. Adaptation of core mechanisms to generate cell polarity. *Nature* **422**, 766–774 (2003).
12. Harris, T. J. & Peifer, M. Adherens junction-dependent and-independent steps in the establishment of epithelial cell polarity in *Drosophila*. *J. Cell Biol.* **167**, 135–147 (2004).
13. Tepass, U. *et al.* shotgun encodes *Drosophila* E-cadherin and is preferentially required during cell rearrangement in the neurectoderm and other morphogenetically active epithelia. *Genes Dev.* **10**, 672–685 (1996).
14. Uemura, T. *et al.* Zygotic *Drosophila* E-cadherin expression is required for processes of dynamic epithelial cell rearrangement in the *Drosophila* embryo. *Genes Dev.* **10**, 659–671 (1996).
15. Cox, R. T., Kirkpatrick, C. & Peifer, M. Armadillo is required for adherens junction assembly, cell polarity, and morphogenesis during *Drosophila* embryogenesis. *J. Cell Biol.* **134**, 133–148 (1996).
16. Peifer, M., Orsulic, S., Sweeton, D. & Wieschaus, E. A role for the *Drosophila* segment polarity gene armadillo in cell adhesion and cytoskeletal integrity during oogenesis. *Development* **118**, 1191–1207 (1993).
17. Petronczki, M. & Knoblich, J. A. DmPAR-6 directs epithelial polarity and asymmetric cell division of neuroblasts in *Drosophila*. *Nature Cell Biol.* **3**, 43–49 (2001).
18. Muller, H. A. & Wieschaus, E. *armadillo*, *bazooka*, and *stardust* are critical for early stages in formation of the zonula adherens and maintenance of the polarized blastoderm epithelium in *Drosophila*. *J. Cell Biol.* **134**, 149–163 (1996).
19. Wodarz, A., Ramrath, A., Grimm, A. & Knust, E. *Drosophila* atypical protein kinase C associates with Bazooka and controls polarity of epithelia and neuroblasts. *J. Cell Biol.* **150**, 1361–1374 (2000).
20. Tepass, U., Theres, C. & Knust, E. crumbs encodes an EGF-like protein expressed on apical membranes of *Drosophila* epithelial cells and required for organization of epithelia. *Cell* **61**, 787–799 (1990).
21. Pilot, F., Philippe, J. M., Lemmers, C., Chauvin, J. P. & Lecuit, T. Developmental control of nuclear morphogenesis and anchoring by charleston, identified in a functional genomic screen of *Drosophila* cellularisation. *Development* **133**, 711–723 (2006).
22. Gabev, E., Kasianowicz, J., Abbott, T. & McLaughlin, S. Binding of neomycin to phosphatidylinositol 4,5-bisphosphate (PIP2). *Biochim Biophys Acta* **979**, 105–112 (1989).
23. Zelhof, A. C. & Hardy, R. W. WASp is required for the correct temporal morphogenesis of rhabdomere microvilli. *J. Cell Biol.* **164**, 417–426 (2004).
24. Formstecher, E. *et al.* Protein interaction mapping: a *Drosophila* case study. *Genome Res.* **15**, 376–384 (2005).
25. Speck, O., Hughes, S. C., Noren, N. K., Kulikauskas, R. M. & Fehon, R. G. Moesin functions antagonistically to the Rho pathway to maintain epithelial integrity. *Nature* **421**, 83–87 (2003).
26. Yonemura, S., Matsui, T., Tsukita, S. & Tsukita, S. Rho-dependent and-independent activation mechanisms of ezrin/radixin/moesin proteins: an essential role for polyphosphoinositides *in vivo*. *J. Cell Sci.* **115**, 2569–2580 (2002).
27. Polesello, C., Delon, I., Valenti, P., Ferrer, P. & Payre, F. Moesin controls actin-based cell shape and polarity during *Drosophila melanogaster* oogenesis. *Nature Cell Biol.* **4**, 782–789 (2002).
28. De Jossineau, C. *et al.* Delta-promoted filopodia mediate long-range lateral inhibition in *Drosophila*. *Nature* **426**, 555–559 (2003).
29. Pinal, N. *et al.* Regulated and polarized PtdIns(3,4,5)P<sub>3</sub> accumulation is essential for apical membrane morphogenesis in photoreceptor epithelial cells. *Curr. Biol.* **16**, 140–149 (2006).
30. von Stein, W., Ramrath, A., Grimm, A., Muller-Borg, M. & Wodarz, A. Direct association of Bazooka/PAR-3 with the lipid phosphatase PTEN reveals a link between the PAR/aPKC complex and phosphoinositide signaling. *Development* **132**, 1675–1686 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank all those who gave us reagents, in particular J. Serano and G. Rubin for fly strains and plasmids, A. Wodarz, A. Zelhof, R. Fehon, F. Payre, E. Suzuki, M. Peifer, H. Oda for flies, antibodies or plasmids; J. Großhans and R. Paro for reagents for synthesizing dsRNA probes for the RNAi screen. We also thank J. Knoblich for suggesting the PIPstrip experiment, members of our group for discussions, S. Kerridge and A. Lebivic for comments on the manuscript. This work was supported by the Association pour la Recherche contre le Cancer (ARC, subvention libre 5179), the CNRS, the Fondation pour la Recherche Médicale (FRM), the Fondation Schlumberger pour l'Education et la Recherche (FSER) and the EMBO Young Investigator Programme. F.P. was supported by the CNRS (bourse BDI) and by the Académie de médecine.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.L. ([lecuit@ibdm.univ-mrs.fr](mailto:lecuit@ibdm.univ-mrs.fr)).

# The physical basis of how prion conformations determine strain phenotypes

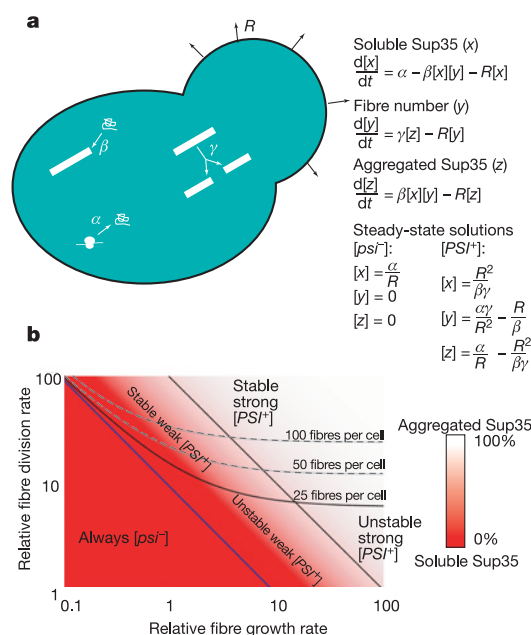
Motomasa Tanaka<sup>1,2</sup>, Sean R. Collins<sup>1</sup>, Brandon H. Toyama<sup>1</sup> & Jonathan S. Weissman<sup>1</sup>

A principle that has emerged from studies of protein aggregation is that proteins typically can misfold into a range of different aggregated forms. Moreover, the phenotypic and pathological consequences of protein aggregation depend critically on the specific misfolded form<sup>1,2</sup>. A striking example of this is the prion strain phenomenon, in which prion particles composed of the same protein cause distinct heritable states<sup>3</sup>. Accumulating evidence from yeast prions such as  $[PSI^+]$  and mammalian prions argues that differences in the prion conformation underlie prion strain variants<sup>3–7</sup>. Nonetheless, it remains poorly understood why changes in the conformation of misfolded proteins alter their physiological effects. Here we present and experimentally validate an analytical model describing how  $[PSI^+]$  strain phenotypes arise from the dynamic interaction among the effects of prion dilution, competition for a limited pool of soluble protein, and conformation-dependent differences in prion growth and division rates. Analysis of three distinct prion conformations of yeast Sup35 (the  $[PSI^+]$  protein determinant) and their *in vivo* phenotypes reveals that the Sup35 amyloid causing the strongest phenotype surprisingly shows the slowest growth. This slow growth, however, is more than compensated for by an increased brittleness that promotes prion division. The propensity of aggregates to undergo breakage, thereby generating new seeds, probably represents a key determinant of their physiological impact for both infectious (prion) and non-infectious amyloids.

The  $[PSI^+]$  phenotype results from self-propagating  $\beta$ -sheet-rich aggregates (amyloid) of the yeast Sup35 translation termination factor<sup>8–10</sup>. Inactivation of Sup35 causes a nonsense suppression phenotype in  $[PSI^+]$  yeast that can be readily monitored using an *ade1* nonsense reporter gene.  $[psi^-]$  colonies require adenine and accumulate a red pigment, whereas  $[PSI^+]$  exhibits a range of heritable strain variants that vary from dark pink to white depending on the degree of Sup35 inactivation<sup>11,12</sup>. Recently, it has been possible to create synthetic prion forms of Sup35 *in vitro* that have adopted distinct infectious (prion) conformations, which when introduced into yeast lead to different prion strain phenotypes<sup>4,5</sup>. This ability to create distinct amyloid conformations of Sup35 *in vitro*, and to monitor their *in vivo* consequences makes  $[PSI^+]$  a uniquely powerful system for investigating what has become a central problem in the field: how do differences in the physical properties of conformationally different misfolded forms modulate their physiological effects?

To pursue this question, we first developed an analytical model that allows us to predict how changes in the physical properties of the Sup35 prion alter observable *in vivo* features of  $[PSI^+]$  such as the fraction of aggregated protein (that is, colour phenotype) and the size and number of the Sup35 aggregates. This effort relies on a number of elegant studies that revealed that prion propagation involves two mechanistically distinct and potentially strain-dependent steps<sup>9,10,13–18</sup>: a growth phase in which unconverted protein adds

onto existing prion particles, and a division phase that generates new prion particles.  $[PSI^+]$  growth results from the self-propagating nature of amyloid fibres, whereas division is mediated by the host chaperone machinery, principally by the Hsp104 chaperone<sup>10,18</sup>. The rate of Sup35 fibre growth is linearly proportional to both the concentrations of soluble Sup35 and prion fibres *in vitro*. Moreover, fibre growth can occur after translation and independently of Hsp104 *in vivo*<sup>10,19,20</sup>, and both *in vitro*<sup>21</sup> and *in vivo*<sup>22</sup> (see also Supplementary Data for details), minimal fibre depolymerization occurs on the timescale of cell division. We therefore modelled fibre growth as a second order irreversible process with a rate constant ( $\beta$ ) that varies



**Figure 1 | An analytical model describing prion strains.** **a**, Left: graphical representation of parameters defining prion dynamics.  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $R$  are the rate constants for translation of Sup35, fibre growth, fibre division and growth of yeast cells, respectively. Right: differential equations describing time-dependent changes in concentration of soluble Sup35 monomer  $[x]$ , Sup35 fibre  $[y]$  and aggregated Sup35  $[z]$ . The two steady-state solutions representing  $[psi^-]$  and  $[PSI^+]$  states are defined by the equations shown. **b**, Graphical representation of the effect of growth and division rates on strain phenotypes. Aggregation states of Sup35 are represented by a colour gradient. Curved lines demark states that have the same number of fibres per cell. The blue line represents a threshold below which stable prion propagation is not possible. Black lines illustrate borders between  $[PSI^+]$  states showing strong or weak colour phenotypes or large (stable) versus small (unstable) numbers of fibres per cell, as indicated.

<sup>1</sup>Howard Hughes Medical Institute, Department of Cellular and Molecular Pharmacology, University of California-San Francisco and California Institute for Quantitative Biomedical Research, San Francisco, California 94143, USA. <sup>2</sup>PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan.

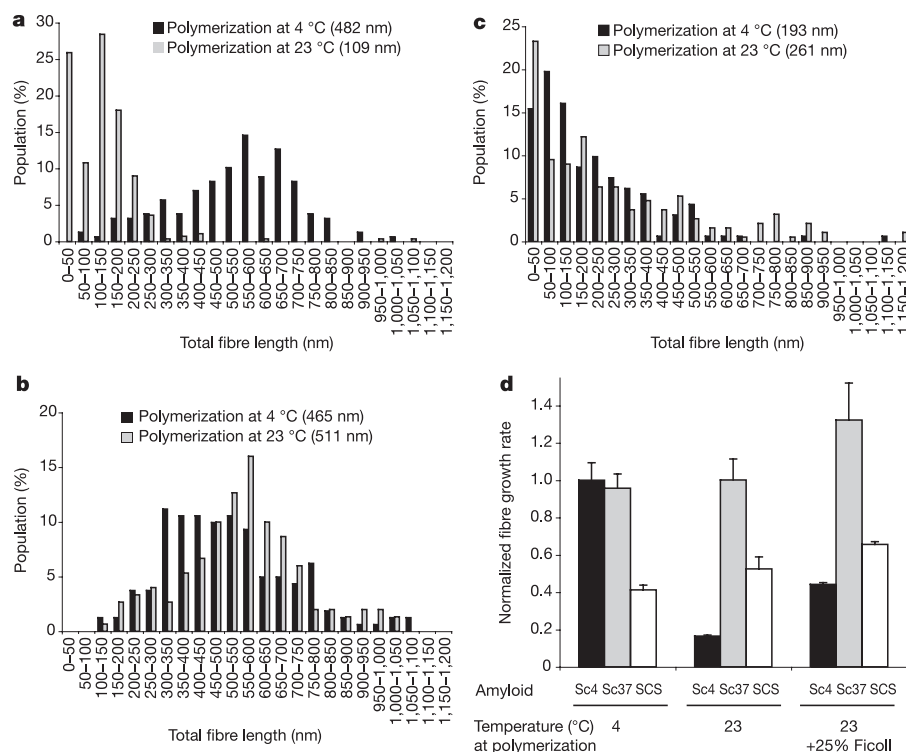
depending on the specific prion strain. The rate of prion replication is proportional to the mass of polymerized Sup35 and this rate of division ( $\gamma$ ) may depend on the prion strain conformation<sup>10,15</sup>. With these considerations in mind, and also accounting for the synthesis of new Sup35 and the dilution of all components through cellular growth, we obtained a set of differential equations that describe the time evolution of the concentration of soluble Sup35 ( $[x]$ ), aggregated Sup35 ( $[z]$ ) and prion particles ( $[y]$ ) in terms of rate constants for translation ( $\alpha$ ), fibre growth ( $\beta$ ), fibre division ( $\gamma$ ) and cellular growth ( $R$ ) (Fig. 1a). One feature not specifically included in this model is lateral association of fibres, as this is expected to have a minimal effect on the underlying processes of fibre growth and division, although lateral association will decrease mitotic stability by reducing the number of independently heritable elements ('propagons') (see Supplementary Data for a more complete discussion).

An important aspect of this analysis, which extends upon previous efforts to model prion dynamics<sup>23–25</sup>, is that it accounts for the existence of distinct stable prion strains even under equilibrium conditions. Specifically, these equations yield two steady-state solutions, one corresponding to a  $[psi^-]$  state and another to a  $[PSI^+]$  state (Fig. 1a). The nature of the  $[PSI^+]$  solution varies with both  $\beta$  and  $\gamma$ , allowing us to predict how the colour phenotype and number of prion fibres depend on the underlying fibre properties (Fig. 1b). The  $[PSI^+]$  solution is non-negative (and hence physically relevant) only for sufficiently great fibre growth and division rates ( $\beta\gamma > R^3/\alpha$ ), but in these cases it is mathematically stable: perturbations altering the concentration of fibres or soluble Sup35 levels are corrected by restoring forces that bring these concentrations back to their steady-state levels. The existence of a continuum of stable steady-state solutions parameterized by the conformation-dependent properties  $\beta$  and  $\gamma$  accounts for how it is possible, even in dividing cells, to have

a broad range of prion strains with differing levels of soluble prion protein and characteristic numbers of fibres<sup>11,12,24</sup> (Fig. 1b). Interestingly, the  $[psi^-]$  solution is not stable and thus is able to persist only because of the extremely high kinetic barrier to *de novo* prion formation. The kinetic instability of the  $[psi^-]$  state helps explain the high efficiency of transmission of  $[PSI^+]$  because, as we confirm by examining the concentration dependence of the efficiency of prion infectivity (Supplementary Fig. S1), introduction of a single prion particle can convert  $[psi^-]$  yeast into  $[PSI^+]$ . Finally, extension of our analysis to explore the phenomenon of strain competition, which occurs when two different conformations are introduced into a cell, indicates that, as previously noted<sup>14</sup>, the prion strain with the lowest levels of soluble protein (highest  $\beta\gamma$  value) is dominant.

We next used our ability to create *in vitro* three distinct infectious conformations of the prion-forming domain (Sup-NM; residues 1–254) of Sup35 to investigate experimentally how changes in the physical properties of an aggregate dictate its resulting *in vivo* strain phenotypes. These three amyloid forms, termed Sc4, Sc37 and SCS, are generated by spontaneously polymerizing Sup-NM at 4 °C or 37 °C, or by seeding polymerization of *Saccharomyces cerevisiae* Sup-NM with seeds derived from the highly divergent *Candida albicans* Sup35 homologue, respectively<sup>26</sup>. Previous studies demonstrated that Sc4, Sc37 and SCS amyloids have well-defined localized conformational differences that can be robustly propagated *in vivo*<sup>5,17,26</sup>. Moreover, infection of  $[psi^-]$  yeast with the Sc4, Sc37 and SCS Sup-NM amyloids leads to readily distinguishable prion strains termed  $[PSI^+(\text{Sc4})]$ ,  $[PSI^+(\text{Sc37})]$  and  $[PSI^+(\text{SCS})]$  that qualitatively show strong (white), intermediate (pink) and weak (dark pink) colour phenotypes, respectively (Supplementary Fig. S2).

We measured the intrinsic fibre growth rate ( $\beta$ ) for these distinct strain conformations of Sup-NM amyloid using an atomic force

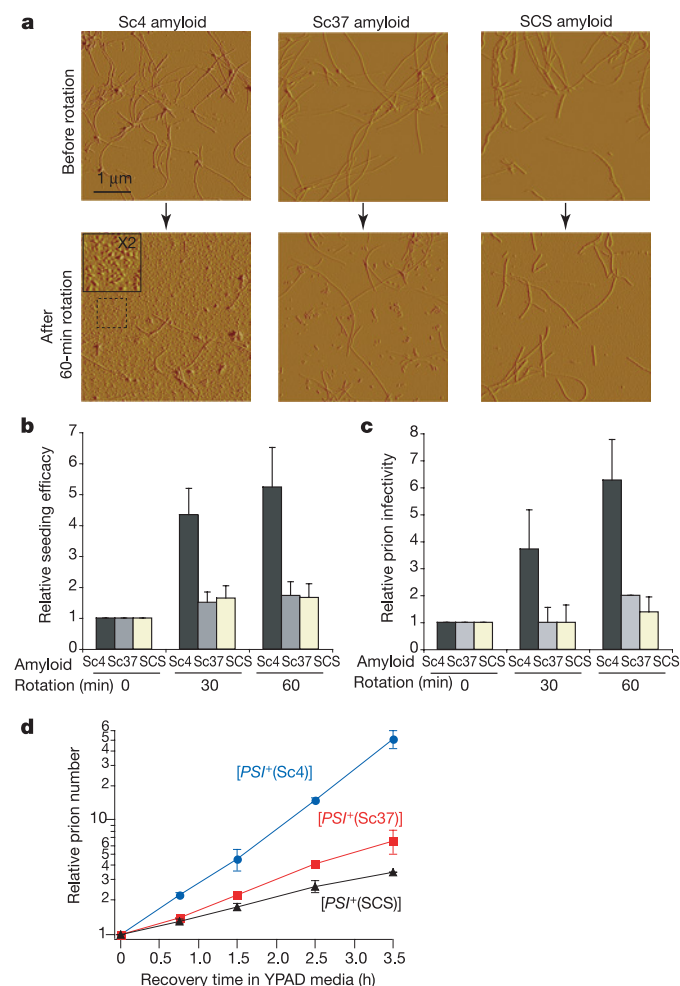


**Figure 2 | Effects of strain conformation on fibre growth rate.** **a–c**, Growth distribution for **(a)** Sc4, **(b)** Sc37 and **(c)** SCS amyloids at 4 °C (black) and 23 °C (grey) (distribution at 37 °C was similar to that at 23 °C, data not shown). Fibre growth rates were determined by an AFM-based single fibre assay<sup>21</sup>. The values indicate mean total fibre length grown over 10 min at 2.5  $\mu$ M soluble Sup-NM. **d**, Bulk measurements of the relative growth rate of Sc4, Sc37 and SCS Sup-NM amyloids. Polymerization of Sup-NM with

indicated fibre seeds was performed at 4 °C or 23 °C in the absence or presence of 25% (wt/wt) Ficoll and the rate of addition of Sup-NM monomers was monitored by thioflavin T fluorescence<sup>26</sup>. The fibre growth rates of each Sup-NM amyloid conformation were normalized with the AFM-based mean length of the corresponding fibres at 4 °C. Values are expressed as mean  $\pm$  s.d.



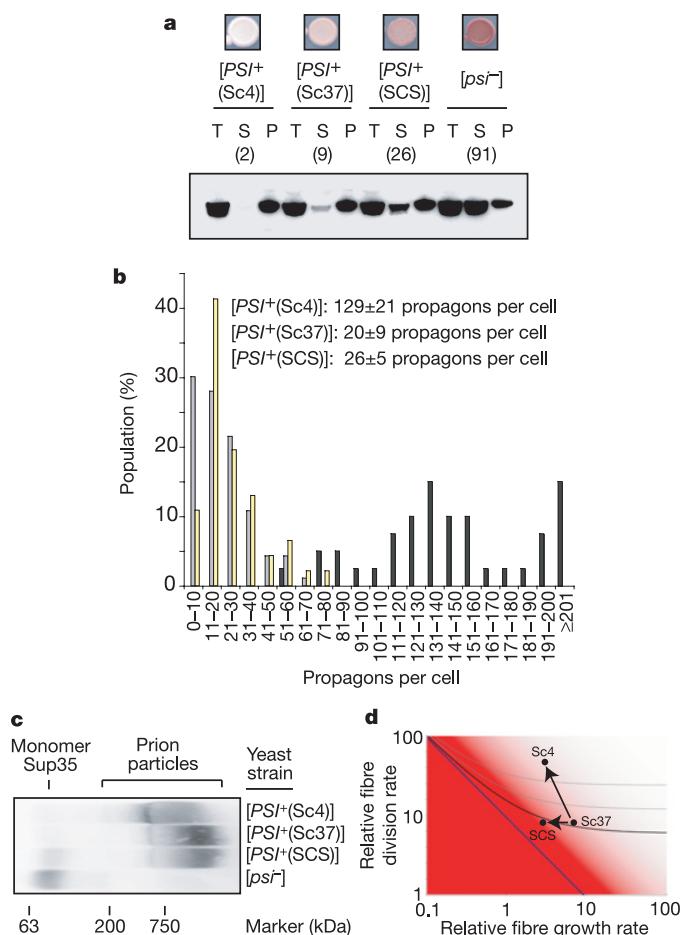
microscopy (AFM)-based single-fibre growth assay<sup>21</sup> (Fig. 2a–c). At physiological temperatures (room temperature and above), the Sc4 fibres showed the slowest growth (109 nm) compared with Sc37 (511 nm) and SCS (261 nm) fibres. Intriguingly, the relative growth rate of Sc4 fibres increases significantly at low temperature (4 °C) (Fig. 2a). Comparison of histograms of fibre growth measurements at the different temperatures reveals that the three conformations have minimally overlapping distributions, arguing strongly that Sc4, Sc37 and SCS fibres represent distinct and conformationally pure forms. We confirmed the slow fibre growth rate of Sc4 amyloid by a bulk seeding assay in the presence of 25% Ficoll, which mimics the macromolecular crowding effects found in cells (Fig. 2d). Thus, we conclude that the Sc4 fibre form shows the slowest intrinsic growth



**Figure 3 | Effects of strain conformation on fibre fragmentation.** **a**, Typical AFM images of Sc4 (left), Sc37 (middle) and SCS (right) Sup-NM amyloids before (top) and after (bottom) stirring for 60 min. Prior to stirring, for all three conformations, most fibres are long (>1  $\mu$ m). After stirring, more than 98% of Sc4 amyloids are <100 nm whereas more than 60% of both Sc37 and SCS amyloids are >100 nm (see Supplementary Fig. S3). A magnified view (black box) of small fibre fragments (dotted box) is also shown for Sc4 fibres. **b**, Relative seeding efficacy<sup>26</sup> of Sc4, Sc37 and SCS Sup-NM fibres before and after stirring for 30 or 60 min. **c**, Relative prion infection efficiency of Sc4, Sc37 and SCS Sup-NM fibres (1  $\mu$ M Sup-NM monomer) before and after rotation for 30 or 60 min. **d**, Effects of strain conformation on prion growth and division *in vivo*. Rate of generation of equilibrium prion state after Gdn-HCl treatment as a function of recovery time for [PSI<sup>+</sup>(Sc4)], [PSI<sup>+</sup>(Sc37)] and [PSI<sup>+</sup>(SCS)] strains. The doubling times of the prions are 37  $\pm$  2, 79  $\pm$  10 and 116  $\pm$  1 min for [PSI<sup>+</sup>(Sc4)], [PSI<sup>+</sup>(Sc37)] and [PSI<sup>+</sup>(SCS)] strains, respectively, which correspond to relative  $\beta\gamma$  values of 3, 1 and 0.67 (see Supplementary Fig. S4 for details). Values are expressed as mean  $\pm$  s.d.

rate (~30% relative to Sc37 in crowding conditions) despite the fact that it leads to the strongest strain phenotype.

Examination of the physical strength of the different prion conformations suggests a solution to this paradox: the slower growth of Sc4 amyloid is accompanied by a marked increase in the propensity of these fibres to be fragmented. We measured the frangibility of the different Sup-NM amyloid conformations *in vitro* by forming long (>1  $\mu$ m) Sup-NM amyloids in the absence of shear forces and then subjecting them to fragmentation by stirring the solution with a magnetic stir bar. AFM analysis revealed that this treatment shattered Sc4 amyloids into many tiny fibres (<100 nm) but had a relatively



**Figure 4 | Analysis of distinct [PSI<sup>+</sup>] strain phenotypes.** **a**, Sedimentation analysis of strains by ultracentrifugation. Total (T), supernatant (S) and pellet (P) fractions of yeast extracts of the indicated strains were analysed by SDS-PAGE followed by western blotting using a polyclonal anti-Sup-NM antibody<sup>30</sup>. Shown in parenthesis are the mean band intensities of supernatant fractions, relative to that of total fractions (100). Typical colour phenotypes of [PSI<sup>+</sup>(Sc4)], [PSI<sup>+</sup>(Sc37)], [PSI<sup>+</sup>(SCS)] and [psi<sup>-</sup>] strains are shown in the top panels. **b**, Frequency of propagon numbers per cell for [PSI<sup>+</sup>(Sc4)] (black), [PSI<sup>+</sup>(Sc37)] (grey) and [PSI<sup>+</sup>(SCS)] (yellow) strains. (See also Supplementary Information.) **c**, Agarose gel analysis<sup>15</sup> of prion particle size in lysates of [PSI<sup>+</sup>(Sc4)], [PSI<sup>+</sup>(Sc37)], [PSI<sup>+</sup>(SCS)] and [psi<sup>-</sup>] strains. Migration positions of prion particles and monomeric Sup35 are indicated. **d**, Strain phenotypes predicted by our model (Fig. 1b) based on the observed growth and division properties of the various [PSI<sup>+</sup>] strains. SCS and Sc4 both show decreased growth rates (~30–50%) compared with Sc37. In the case of SCS, this is not accompanied by an increase in division rate and thus leads to the observed increase in soluble Sup35. In contrast, for Sc4 the large increase in the division rate offsets the slow growth ( $\beta\gamma$  increases ~300% relative to Sc37), leading to a decreased pool of Sup35 and an even more dramatic increase in propagon number (decrease in size). As in Fig. 1b, the calculated amount of soluble Sup35 is shown with a red–white scale.

modest effect on Sc37 and SCS fibres (Fig. 3a and Supplementary Fig. S3). Concurrent with the appearance of new fibre ends, we observed a marked increase in seeding efficacy and prion infectivity for the Sc4 fibres but not the Sc37 or SCS fibres (Fig. 3b, c). The increase in fragility of the Sc4 fibres is consistent with their increased sensitivity to denaturants and proteases relative to Sc37 and SCS fibres<sup>5,17,26</sup> and probably results from strain-specific structural differences in inter molecular contacts and a shorter, less stable amyloid core<sup>17,26</sup>.

We next examined whether Sc4 also shows an increased rate of division under physiological conditions in intact cells, using an elegant approach developed previously<sup>18,22</sup>. Specifically, the rate at which the equilibrium prion state is restored after transient growth in guanidine hydrochloride (Gdn-HCl)-containing media, which reversibly inhibits Hsp104 (and thus prion division)<sup>27,28</sup>, reports on the product of the intrinsic propensity of fibres to be divided per unit length ( $\gamma$ ) and the growth rate ( $\beta$ ) (see Supplementary Fig. S4). The [ $PSI^+$ (Sc4)] strain regenerated the equilibrium prion state after removal of Gdn-HCl substantially more rapidly than [ $PSI^+$ (Sc37)], which itself was faster than the [ $PSI^+$ (SCS)] strain (Fig. 3d). Thus, the slow growth rate (low  $\beta$ ) relative to Sc37 fibre amyloid is more than overcome by the faster *in vivo* prion division rate for the Sc4 but not the SCS strain.

We used a range of assays to define the *in vivo* phenotypes of the [ $PSI^+$ (Sc4)], [ $PSI^+$ (Sc37)] and [ $PSI^+$ (SCS)] strains. Subjecting extracts to high-speed centrifugation revealed that, as would be expected from its weak colour phenotype, the [ $PSI^+$ (SCS)] strain contains a substantial pool of soluble Sup35, compared to a modest one for [ $PSI^+$ (Sc37)] and a barely detectable one for [ $PSI^+$ (Sc4)] (Fig. 4a). In contrast, examination of prion number and size by several independent assays indicated that, as expected by its relative fragility, [ $PSI^+$ (Sc4)] is distinguished from the other two strains by having a large number of relatively small prion elements. A 'propagion counting' assay<sup>18</sup> (see also Supplementary Data for details) indicated that the apparent number of propagons per cell in the [ $PSI^+$ (Sc4)] strain was much larger ( $129 \pm 21$ ) than that found in either [ $PSI^+$ (Sc37)] ( $20 \pm 9$ ) or [ $PSI^+$ (SCS)] ( $26 \pm 5$ ) (Fig. 4b). Furthermore, both agarose gel analysis in the presence of SDS (Fig. 4c), which is thought to report on the length of individual prion fibres<sup>15</sup>, and sedimentation of prion particles in non-denaturing conditions (Supplementary Fig. S5), established that the prion particles in the [ $PSI^+$ (Sc4)] strain are smaller than those found in either [ $PSI^+$ (Sc37)] or [ $PSI^+$ (SCS)] strains.

As illustrated in Fig. 4d, the above strain phenotypes are in excellent agreement with those predicted by our analysis based on their observed prion growth and division properties. In particular, the slow growth of SCS amyloids (~30–50% relative to Sc37), which is not accompanied by an increase in division rate, accounts for the observed increase in the level of soluble Sup35 and the correspondingly weaker colour phenotype. In contrast, although Sc4 also grows more slowly than Sc37, this is over-ridden by a much higher division rate ( $\beta\gamma$  is estimated to increase ~3-fold, see Supplementary Data for details), leading to a decrease in the amount of soluble Sup35 (that is, white colour phenotype) and an even more dramatic increase in the number (and thus decrease in size) of fibres.

Our work establishes the critical role of the frangibility of a prion aggregate in dictating the strength of its strain phenotype. Less is known for other systems about the mechanism of amyloid growth and division. Nonetheless, similar dynamic effects, with dilution mediated by proteolysis and clearance rather than cell division in post-mitotic cells, probably apply for other aggregation phenomena including both infectious (prion) and non-infectious amyloids<sup>23–25</sup>. Indeed, recent studies indicate that smaller aggregates of the mammalian prion protein PrP, which should be more readily generated by strains that form fragile particles, are markedly more infectious than larger aggregates<sup>29</sup>. Additionally, for synthetic mammalian prions, the incubation time increases as the physical stability of the prion increases (G. Legname, H. O. Nguyen, D. Peretz, F. E. Cohen, S. J. DeArmond and

S. B. Prusiner, personal communication). Finally, it has been observed that stronger [ $PSI^+$ ] strains typically have shorter fibres *in vivo* than weaker strains<sup>15</sup>, although the physical properties of the prions (for example, their propensity to be fragmented) were not examined. Taken together, these considerations argue that variability in the brittleness of aggregates represents a major mechanism by which prion conformations dictate the strength of strain phenotypes. Thus, a strong impediment to the production of highly infectious synthetic mammalian prions may be the propensity to generate highly stable amyloid forms *in vitro*<sup>6,16</sup>. Conversely, therapeutic efforts to minimize the impact of protein aggregation could be tailored not only to slow aggregate formation and growth but also to decrease the rate of fragmentation either by increasing fibre stability or inhibiting chaperone machinery. More broadly, this analysis allows for a unified analytical treatment of both non-infectious and infectious (prion) amyloids with prion states arising only when the rate of amyloid growth, division and spread is sufficient to allow sustained propagation.

## METHODS

**Yeast strains and fibre preparation.** We used isogenic [ $psi^-$ ][ $RNQ^+$ ] and [ $PSI^+$ ] derivatives of the 74D-694 yeast strain<sup>30</sup>. [ $PSI^+$ (Sc4)], [ $PSI^+$ (Sc37)] and [ $PSI^+$ (SCS)] strains were made by infection of [ $psi^-$ ] yeast with *in-vitro*-produced Sup-NM Sc4, Sc37 and SCS amyloids<sup>26</sup>. Sc4, Sc37 and SCS (previously termed as Sc[Ca<sub>3</sub>Sc4]) fibres were produced as described previously<sup>26</sup>, using bacterially produced pure Sup-NM proteins carboxy-terminally tagged with 7×histidine<sup>25</sup>.

***In vitro* analysis of strain conformations of Sup-NM amyloids.** The AFM-based single-fibre assay was performed as reported previously<sup>21</sup> with some modification. Each histogram involved the measurement of at least 150 individual fibres. The fibre growth rate of distinct Sup-NM amyloid fibres was also examined by thioflavin T fluorescence<sup>26</sup>, in the absence or presence of Ficoll PM70 to the final concentration of 25% (wt/v). For fibre rigidity assay, Sup-NM amyloids were formed by polymerizing Sup-NM (5  $\mu$ M) in a beaker (2.2 cm diameter), containing 2 ml of 5 mM potassium phosphate buffer containing 150 mM NaCl (pH 7.4), overnight under undisturbed conditions in the presence of 5% (mol/mol) seed of Sc4, Sc37 and SCS amyloids at 4, 37 and 23 °C, respectively. The amyloid solution was stirred by a magnetic stir bar (3 mm diameter, 1 cm length) at ~100 r.p.m. and an aliquot was taken at 0, 30 and 60 min and used for morphological, seeding efficacy<sup>26</sup> and prion infectivity<sup>5</sup> analyses. The detailed methods are described in Supplementary Data.

***In vivo* analysis of prion strains.** The curing assay of [ $PSI^+$ ] with Gdn-HCl was followed by a previously developed procedure<sup>18,22</sup>. The propagion counting experiment was performed as described previously<sup>18</sup>. For agarose gel analysis, yeast cell lysates were prepared with glass beads, and prion particles in the lysates were separated by horizontal 1.6% agarose gels in TAE buffer containing 0.1% SDS<sup>15</sup>, followed by immunoblotting with a polyclonal anti-Sup-NM antibody<sup>30</sup>. Sedimentation analysis was performed as reported previously<sup>5</sup> with some modification. The detailed methods are described in Supplementary Data.

Received 27 February; accepted 3 May 2006.

Published online 28 June 2006.

1. Caughey, B. & Lansbury, P. T. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26**, 267–298 (2003).
2. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
3. Collinge, J. Prion diseases of humans and animals: their causes and molecular basis. *Annu. Rev. Neurosci.* **24**, 519–550 (2001).
4. King, C. Y. & Diaz-Avalos, R. Protein-only transmission of three yeast prion strains. *Nature* **428**, 319–323 (2004).
5. Tanaka, M., Chien, P., Naber, N., Cooke, R. & Weissman, J. S. Conformational variations in an infectious protein determine prion strain differences. *Nature* **428**, 323–328 (2004).
6. Legname, G. *et al.* Synthetic mammalian prions. *Science* **305**, 673–676 (2004).
7. Brachmann, A., Baxa, U. & Wickner, R. B. Prion generation *in vitro*: amyloid of Ure2p is infectious. *EMBO J.* **24**, 3082–3092 (2005).
8. Wickner, R. B. [URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science* **264**, 566–569 (1994).
9. Tuite, M. F. & Koloteva-Levin, N. Propagating prions in fungi and mammals. *Mol. Cell* **14**, 541–552 (2004).
10. Shorter, J. & Lindquist, S. Prions as adaptive conduits of memory and inheritance. *Nature Rev. Genet.* **6**, 435–450 (2005).
11. Derkatch, I. L., Chernoff, Y. O., Kushnirov, V. V., Inge-Vechtomov, S. G. & Liebman, S. W. Genesis and variability of [ $PSI^+$ ] prion factors in *Saccharomyces cerevisiae*. *Genetics* **144**, 1375–1386 (1996).

12. Kochneva-Pervukhova, N. V. *et al.*  $[PSI^+]$  prion generation in yeast: characterization of the 'strain' difference. *Yeast* **18**, 489–497 (2001).
13. Kushnirov, V. V. & Ter-Avanesyan, M. D. Structure and replication of yeast prions. *Cell* **94**, 13–16 (1998).
14. Bradley, M. E., Edskes, H. K., Hong, J. Y., Wickner, R. B. & Liebman, S. W. Interactions among prions and prion "strains" in yeast. *Proc. Natl Acad. Sci. USA* **99** (suppl. 4), 16392–16399 (2002).
15. Kryndushkin, D. S., Alexandrov, I. M., Ter-Avanesyan, M. D. & Kushnirov, V. V. Yeast  $[PSI^+]$  prion aggregates are formed by small Sup35 polymers fragmented by Hsp104. *J. Biol. Chem.* **278**, 49636–49643 (2003).
16. Castilla, J., Saa, P., Hetz, C. & Soto, C. *In vitro* generation of infectious scrapie prions. *Cell* **121**, 195–206 (2005).
17. Krishnan, R. & Lindquist, S. L. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature* **435**, 765–772 (2005).
18. Cox, B., Ness, F. & Tuite, M. Analysis of the generation and segregation of propagons: entities that propagate the  $[PSI^+]$  prion in yeast. *Genetics* **165**, 23–33 (2003).
19. Collins, S. R., Douglass, A., Vale, R. D. & Weissman, J. S. Mechanism of prion propagation: amyloid growth occurs by monomer addition. *PLoS Biol.* **2**, e321 (2004).
20. Satpute-Krishnan, P. & Serio, T. R. Prion protein remodelling confers an immediate phenotypic switch. *Nature* **437**, 262–265 (2005).
21. DePace, A. H. & Weissman, J. S. Origins and kinetic consequences of diversity in Sup35 yeast prion fibers. *Nature Struct. Biol.* **9**, 389–396 (2002).
22. Ness, F., Ferreira, P., Cox, B. S. & Tuite, M. F. Guanidine hydrochloride inhibits the generation of prion "seeds" but not prion protein aggregation in yeast. *Mol. Cell. Biol.* **22**, 5593–5605 (2002).
23. Masel, J., Jansen, V. A. & Nowak, M. A. Quantifying the kinetic parameters of prion replication. *Biophys. Chem.* **77**, 139–152 (1999).
24. Weissmann, C. The state of the prion. *Nature Rev. Microbiol.* **2**, 861–871 (2004).
25. Hall, D. & Edskes, H. Silent prions lying in wait: a two-hit model of prion/amyloid formation and infection. *J. Mol. Biol.* **336**, 775–786 (2004).
26. Tanaka, M., Chien, P., Yonekura, K. & Weissman, J. S. Mechanism of cross-species prion transmission: an infectious conformation compatible with two highly divergent yeast prion proteins. *Cell* **121**, 49–62 (2005).
27. Ferreira, P. C., Ness, F., Edwards, S. R., Cox, B. S. & Tuite, M. F. The elimination of the yeast  $[PSI^+]$  prion by guanidine hydrochloride is the result of Hsp104 inactivation. *Mol. Microbiol.* **40**, 1357–1369 (2001).
28. Jung, G. & Masison, D. C. Guanidine hydrochloride inhibits Hsp104 activity *in vivo*: a possible explanation for its effect in curing yeast prions. *Curr. Microbiol.* **43**, 7–10 (2001).
29. Silveira, J. R. *et al.* The most infectious prion protein particles. *Nature* **437**, 257–261 (2005).
30. Santoso, A., Chien, P., Osherovich, L. Z. & Weissman, J. S. Molecular basis of a yeast prion species barrier. *Cell* **100**, 277–288 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank G. Legname and S. Prusiner for communicating their results before publication, B. Cox and T. Serio for personal communication, and T. C. Keller III for providing us with chicken pectoralis extracts. We also thank C. Cunningham, J. Newman, L. Osherovich, M. Schuldiner, K. Tipton and members of the Weissman laboratory for helpful discussion and critical reading of the manuscript. M.T. was partly supported by JSPS and Uehara Memorial postdoctoral fellowships for research abroad. S.R.C. was supported by predoctoral fellowships from the Burroughs Wellcome Fund. Funding was also provided by the Howard Hughes Medical Institute, The David and Lucile Packard Foundation and the National Institutes of Health (J.S.W.).

**Author Contributions** S.R.C. led the development of the analytical model. M.T. was responsible for the execution of the experiments, with the exception of the fibre growth studies, which were conducted by B.H.T. and M.T. M.T., S.R.C. and J.S.W. were primarily responsible for the design, interpretation and written description of the results.

**Author Information** Reprints and permission information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for information should be addressed to J.S.W. ([weissman@cmp.ucsf.edu](mailto:weissman@cmp.ucsf.edu)).



# Rad54 protein promotes branch migration of Holliday junctions

Dmitry V. Bugreev<sup>1,2</sup>, Olga M. Mazina<sup>1</sup> & Alexander V. Mazin<sup>1</sup>

Homologous recombination has a crucial function in the repair of DNA double-strand breaks and in faithful chromosome segregation<sup>1–3</sup>. The mechanism of homologous recombination involves the search for homology and invasion of the ends of a broken DNA molecule into homologous duplex DNA to form a cross-stranded structure, a Holliday junction (HJ)<sup>4–7</sup>. A HJ is able to undergo branch migration along DNA, generating increasing or decreasing lengths of heteroduplex. In both prokaryotes and eukaryotes, the physical evidence for HJs, the key intermediate in homologous recombination, was provided by electron microscopy<sup>8</sup>. In bacteria there are specialized enzymes that promote branch migration of HJs<sup>7</sup>. However, in eukaryotes the identity of homologous recombination branch-migration protein(s) has remained elusive. Here we show that Rad54, a Swi2/Snf2 protein<sup>9</sup>, binds HJ-like structures with high specificity and promotes their bidirectional branch migration in an ATPase-dependent manner. The activity seemed to be conserved in human and yeast Rad54 orthologues. *In vitro*, Rad54 has been shown to stimulate DNA pairing of Rad51, a key homologous recombination protein<sup>10–12</sup>. However, genetic data indicate that Rad54 protein might also act at later stages of homologous recombination, after Rad51 (ref. 13). Novel DNA branch-migration activity is fully consistent with this late homologous recombination function of Rad54 protein.

Rad54 is a member of the Rad52 family of homologous recombination proteins; its inactivation causes a marked deficiency in homologous recombination and DNA repair in all eukaryotes<sup>14,15</sup>. Biochemical studies indicate that Rad54 can translocate along double-stranded DNA (dsDNA) in an ATPase-dependent manner, in a similar manner to that of prokaryotic branch migration (BM)

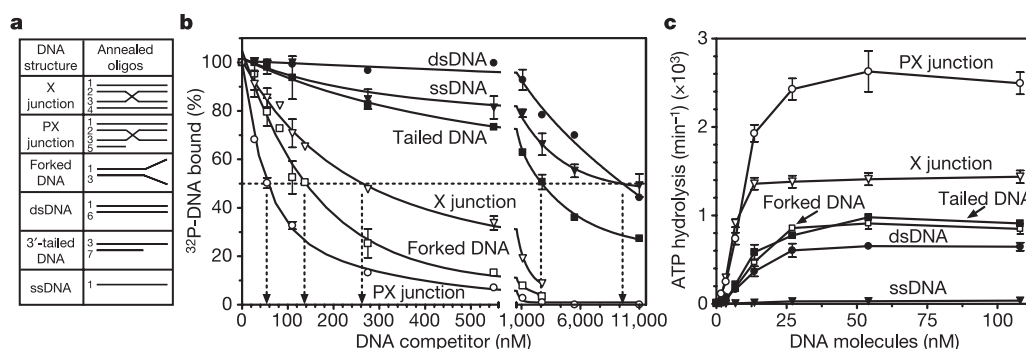
proteins<sup>12,16,17</sup>. We therefore examined whether Rad54 can also promote BM along DNA.

Prokaryotic DNA BM proteins bind specifically to cruciform DNA substrates resembling a HJ<sup>18</sup> (illustrated in Supplementary Fig. 1). Here we examined the DNA binding specificity of human Rad54 (hRad54) with the use of a band-shift assay (Supplementary Fig. 2) and found that it had a strong preference for a PX junction and to a smaller extent for other branched DNA substrates: forked DNA and X junction (Fig. 1a, b). The concentrations of the DNA competitors required for a 50% decrease in amounts of hRad54 complexes formed with a PX junction were the following: 55 nM PX junction, 135 nM forked DNA, 260 nM X junction, 2,700 nM 3'-tailed DNA, and 9,500 nM dsDNA or single-stranded DNA (ssDNA) (Fig. 1b).

We tested then the DNA binding preference of hRad54 in an ATPase assay, because the ATPase activity of hRad54 depends almost entirely on DNA binding<sup>19</sup>. In accord with the results of the gel-shift assay, hRad54 showed a preference for branched DNA, especially for a PX junction (Fig. 1c).

Thus, similarly to known BM proteins, hRad54 shows a strong affinity for branched DNA substrates<sup>18</sup>. The strongest preference is for a PX junction, which structurally resembles one end of a D-loop (Supplementary Fig. 1), the product of the invasion step of homologous recombination—an appropriate substrate for DNA BM proteins.

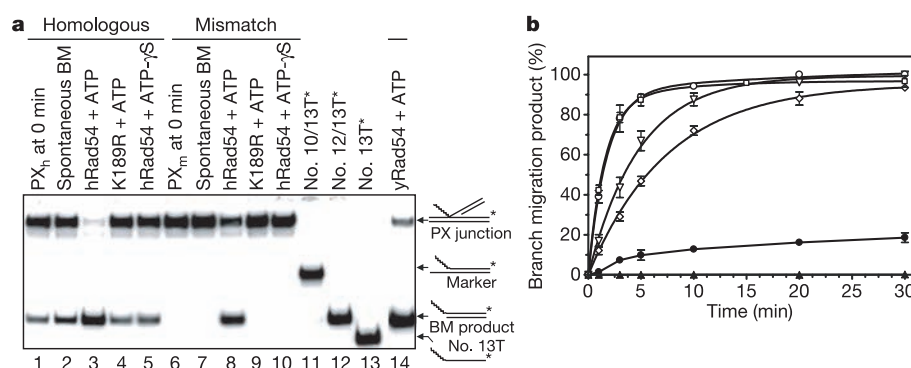
We investigated then whether hRad54 can catalyse the BM of synthetic DNA junctions that mimic a HJ. Two synthetic PX junctions were prepared by annealing two partial duplexes (Supplementary Fig. 3). In the first, the duplex of the invading oligonucleotide was completely homologous to the duplex of the target oligonucleotide, permitting unimpeded BM. In the second PX junction, the



**Figure 1 | hRad54 protein binds preferentially to branched DNA molecules.** **a**, Tested DNA substrates; the sequences are shown in Supplementary Table 1. **b**, Graphical representation of binding of <sup>32</sup>P-labelled PX junction to hRad54 in the presence of the indicated

unlabelled DNA competitors analysed by gel electrophoresis in polyacrylamide gels (see Supplementary Fig. 2). **c**, Effect of DNA substrates on the ATPase activity of hRad54. Error bars indicate s.e.m.

<sup>1</sup>Department of Biochemistry and Molecular Biology, Drexel University College of Medicine, Philadelphia, Pennsylvania 19102-1192, USA. <sup>2</sup>Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of Russian Academy of Science, Novosibirsk, 630090, Russia.



**Figure 2 | Rad54 promotes an ATPase-dependent BM of synthetic PX junction DNA.** The scheme of the BM assay with either fully homologous or mismatched duplexes is shown on Supplementary Fig. 3. **a**, The products of BM revealed by gel electrophoresis. Lanes 1 and 6 show two synthetic PX junctions (PX<sub>h</sub> denotes homologous junction and PX<sub>m</sub> a mismatched one). BM was initiated by the addition of hRad54, yRad54 (lane 14, with PX<sub>m</sub> junction) or the hRad54<sup>K189R</sup> mutant protein to the PX junction DNA and performed for 15 min at 30 °C in the presence of the indicated nucleotide

cofactors. In control reactions, Rad54 protein was replaced by storage buffer (lanes 2 and 7, 'Spontaneous BM'). Lanes 11–13 show the migration markers. The asterisk indicates <sup>32</sup>P label. **b**, The kinetics of BM, spontaneous or promoted by yRad54 or hRad54 protein, on either fully homologous or mismatched PX DNA structures. Open squares, homologous yRad54; open circles, homologous hRad54; open triangles, mismatch yRad54; open diamonds, mismatch hRad54; filled circles, spontaneous homologous; filled triangles, spontaneous mismatch. Data are means ± s.e.m.

sequences of the duplexes of the target and the invading oligonucleotides differed by a single base pair; DNA BM on this substrate would create a mismatch whose bypass would require an energy input<sup>20</sup>.

We found that hRad54 greatly accelerated BM on the fully homologous PX junction, in comparison with spontaneous BM (Fig. 2a, lanes 2 and 3). As expected for BM proteins, ATP hydrolysis seemed to be essential for the BM activity of hRad54; BM was inhibited in the presence of ATP-γS and was not promoted by the ATPase-deficient hRad54K189R mutant (Fig. 2a, lanes 4 and 5).

We addressed the possibility that the BM product (Fig. 2a) could be formed by the melting of DNA substrates followed by cross-reannealing of the resultant ssDNA strands. We showed that the BM product was not produced when the formation of a PX junction, a substrate for BM, was precluded by the incorporation of a heterologous sequence in the ssDNA tail of one of the two DNA substrates (Supplementary Fig. 4).

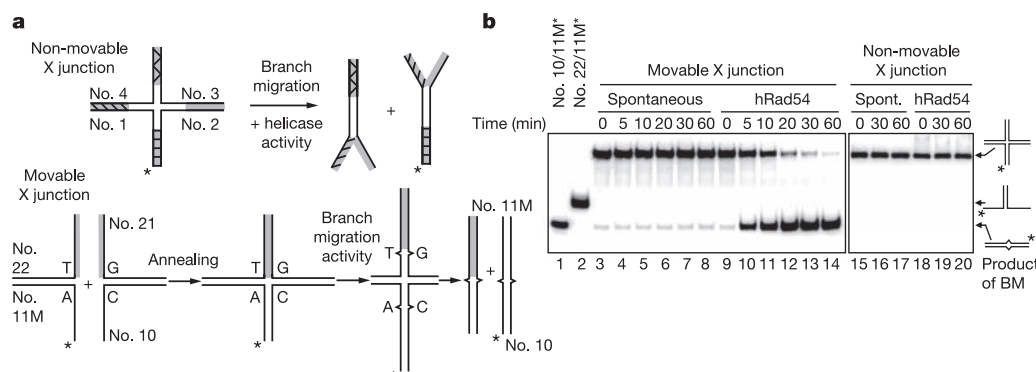
Studies with the PX junction containing a single-base mismatch in duplex DNA branches (Supplementary Fig. 3b) were the most revealing for the BM activity of hRad54. Whereas spontaneous BM was completely blocked (Fig. 2a, lane 7), hRad54 efficiently promoted bypass of the mismatch in an ATPase-dependent manner

(Fig. 2a, lanes 8–10). However, the initial rate of BM was inhibited approximately threefold by the mismatch (Fig. 2b).

We also found that yeast Rad54 (yRad54), similarly to hRad54, efficiently promotes BM on PX junctions, both homologous and mismatched (Fig. 2a, lane 14; Fig. 2b). Thus, the BM activity of Rad54 on DNA seems to be evolutionarily conserved.

We examined whether structural asymmetry of the PX junctions affects the direction of hRad54 BM. It seems that, similarly to the RuvAB protein complex<sup>21</sup>, hRad54 can promote four-stranded BM in either the 3' → 5' or the 5' → 3' direction relative to the displaced ssDNA strand (Supplementary Fig. 5). We also tested whether hRad54 can promote a three-stranded reaction (heteroduplex extension). With a PX junction containing heterologous duplex DNA branches in which four-stranded BM was blocked, we found that hRad54 did indeed efficiently promote three-stranded BM in an ATPase-dependent manner (Supplementary Fig. 6).

We tested further whether hRad54 can promote BM on an X junction, a common substrate for BM proteins. Rad54, in contrast to other known BM proteins, has no DNA helicase activity<sup>10</sup>. Therefore conventional X junctions, in which a central homologous 'movable' core is flanked by the mutually heterologous terminal DNA branches that block BM (ref. 22) (Fig. 3a, top), seemed to be unsuitable for its



**Figure 3 | Rad54 promotes BM of synthetic four-way HJs (X junctions).** **a**, The scheme. The non-movable X junctions (top) require DNA helicase activity for their disruption. The movable X junction (bottom) contained a mismatch to block spontaneous BM. Shaded regions denote heterologous DNA terminal branches. The asterisk indicates <sup>32</sup>P label at the DNA 5' end. **b**, The products of BM were revealed by electrophoresis in an 8% polyacrylamide gel at 4 °C. BM was initiated by addition of hRad54 to X

junctions and performed for the indicated periods at 30 °C (lanes 9–14, hRad54). In control reactions, hRad54 protein was replaced by storage buffer (lanes 3–8, spontaneous). Lines 15–20 show lack of spontaneous (Spont.) or hRad54-dependent (hRad54) BM on the non-movable X junctions. Lanes 1 and 2 show the migration markers. Quantification of the data is shown in Supplementary Fig. 7.

BM activity (Fig. 3b, right); DNA helicase activity, which disrupts the 'non-movable' branches, is needed to detect BM on these substrates. We therefore constructed an X junction in which three of the four terminal DNA branches were mutually homologous, allowing the crossover point to move freely by BM in one direction up to complete separation of two DNA duplexes, without any need for a helicase activity (Fig. 3a, bottom). Using this new substrate we found that hRad54 can efficiently branch-migrate X junctions, similarly to PX junctions, bypassing a mismatched base pair that blocks spontaneous BM (Fig. 3b; Supplementary Fig. 7).

An important question is whether the hRad54 protein can promote BM on DNA structures that are intermediates of DNA strand exchange produced by hRad51 on long DNA substrates. To generate such structures we used a four-stranded DNA exchange reaction (Supplementary Fig. 8b)<sup>23</sup>. The difference between conventional three-stranded (Supplementary Fig. 8a) and four-stranded reactions is significant. The three-stranded reaction includes only DNA heteroduplex formation and its extension; the HJ is not formed in this reaction. In contrast, in the four-stranded DNA reaction between two at least partly double-stranded DNA molecules, the extension leads to formation of the crossover point. BM proteins can move the crossover point along the two helices, extending the region of heteroduplex DNA.

First, using <sup>32</sup>P-labelled linear pBSK (+) dsDNA and the gapped DNA, prepared by annealing the *XhoI*-*AlwNI* fragment (2,065 base pairs (bp)) to circular ssDNA pBSK (+), we examined whether hRad51 alone can promote BM in a four-stranded DNA exchange. We found that hRad51 protein failed to generate any significant amounts of a nicked-circular DNA product in a four-stranded reaction within 2 h, although it efficiently promoted joint molecule formation (Fig. 4a, lanes 5–8). In contrast, in a three-stranded DNA exchange between pBSK (+) ssDNA and linear dsDNA, hRad51 efficiently promoted the formation of both joint molecules and a nicked-circular DNA product (Fig. 4a, lanes 1–4). The accumulation of DNA joint molecule intermediates in a four-stranded reaction was consistent with the formation of  $\alpha$ -structures in the proximity of the ssDNA-dsDNA junction<sup>23</sup>. We also found that under the examined conditions hRad51 is unable to promote four-stranded BM through even a short dsDNA region (114 bp; Supplementary Fig. 9b, c). In contrast, RecA efficiently promoted four-stranded DNA exchange, as expected (Supplementary Fig. 9a).

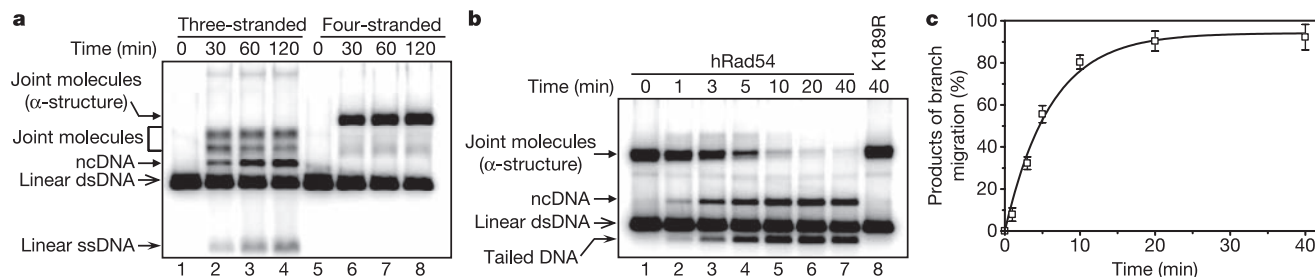
We next tested whether hRad54 protein can promote BM of the DNA intermediates ( $\alpha$ -structures) generated by hRad51 in the four-stranded DNA exchange. When added to the deproteinized DNA intermediates (Fig. 4b, lane 1) hRad54 efficiently promoted BM through a 2,065-bp dsDNA region generating nicked-circular and tailed DNA products (Fig. 4b, lanes 2–7) and completing the reaction in less than 20 min (Fig. 4c). As expected, the ATPase activity of hRad54 was essential for BM (Fig. 4b, lane 8). With non-deproteinized  $\alpha$ -structures we did not see an appearance of the DNA products

(data not shown), which would involve BM through a 2,065-bp dsDNA region covered by hRad51. However, we could not exclude the possibility that limited BM takes place on these substrates. To address this question we used shorter DNA substrates and found that hRad54 protein has an intrinsic ability to promote BM of non-deproteinized DNA intermediates formed by hRad51 (Supplementary Fig. 10), paralleling the ability of the bacterial RuvAB complex to promote BM in the presence of RecA<sup>23</sup>. Apparently, BM of non-deproteinized DNA intermediates involved specific protein–protein interactions, because  $\gamma$ Rad54 promoted it significantly less efficiently (Supplementary Fig. 10b, c).

Similarly to Rad54, ISWI (Imitation SWI) and hSWI/SNF, members of the SWI2/SNF2 family, can translocate along dsDNA in an ATPase-dependent manner<sup>24</sup>. However, we found that under the conditions tested they failed to promote BM of the PX junctions, either with a mismatch (Supplementary Figs 11 and 12) or with a fully homologous junction (data not shown). Thus, BM activity seems to be specific for Rad54 rather than being a general property of the SWI2/SNF2 proteins.

Several eukaryotic helicases, namely WRN, BLM and RECQ5 $\beta$ , can catalyse BM in addition to their conventional unwinding of duplex DNA<sup>7</sup>. However, these enzymes are unlikely to be homologous recombination BM proteins; they disrupt HJs *in vitro* and show anti-recombinogenic characteristics *in vivo*<sup>7,25</sup>. Rad54 seems to be a unique BM protein because, in contrast to other known BM enzymes, it has no DNA helicase activity<sup>10</sup>. Rad54 BM activity might previously have evaded detection<sup>26</sup> because it requires specifically tailored DNA substrates such as those described above. Still, it is likely that other as yet unknown eukaryotic BM proteins exist<sup>27</sup>, whose relationship with Rad54 remains to be determined.

Because of the binding preference of hRad54 for the PX junction, which structurally resembles one end of the D-loops, we propose that D-loops might represent natural substrates for BM activity of hRad54 *in vivo*. Because hRad54 promotes both three-stranded and four-stranded BM, we propose that hRad54 can promote BM of the DNA junctions at the D-loop in either direction, away from or towards the DNA double-strand break (DSB) (Supplementary Fig. 1a). In the first case it stabilizes D-loops, facilitating the capture of the second DNA end and promoting the formation of a double HJ (Supplementary Fig. 1b, left), thereby contributing to the mechanism described by the DSB repair (DSBR) model<sup>2</sup>. In accord with this model, resolution of the double HJ produces both crossover and non-crossover recombinants. Relevant to this model,  $\gamma$ Rad54 was shown to interact physically with the junction-specific endonuclease Mus81/Mms4 (ref. 28). In the second case it facilitates dissociation of the invading DNA end from the template DNA strand, disrupting D-loops (Supplementary Fig. 1b, right), in accord with the synthesis-dependent strand annealing (SDSA) model<sup>2</sup>, which postulates the formation of mostly non-crossover recombinants. Genetic studies in yeast indicate that the Rad54 may contribute to DSB repair



**Figure 4 | hRad54 protein catalyses BM of joint molecules ( $\alpha$ -structures) generated by hRad51 in four-stranded DNA exchange.** The scheme is shown in Supplementary Fig. 8. **a**, DNA strand exchange between <sup>32</sup>P-labelled pBSK linear dsDNA and circular ssDNA (three-stranded) or pBSK gapped DNA (four-stranded) promoted by hRad51. **b**, Joint molecules

( $\alpha$ -structures) generated by hRad51 protein in the four-stranded DNA exchange, shown in **a**, were used for BM assay after DNA deproteinization. BM was initiated by the addition of hRad54 (lanes 2–7) or hRad54<sup>K189R</sup> mutant protein (lane 8). **c**, Graphical representation of the kinetics of BM shown in **b**. Error bars indicate s.e.m.



mechanisms described by both the DSBR and SDSA models, and the contribution varies from one experimental system to another<sup>13,29</sup>.

Although Rad54 protein is likely to participate in homologous recombination at different stages<sup>3</sup> and promote chromatin remodeling<sup>24</sup>, genetic data in yeast indicate the significance of its postsynaptic function, downstream of the Rad51 protein<sup>13</sup>. The previously unknown DNA BM activity that we have identified and characterized is consistent with the postsynaptic function of Rad54 in homologous recombination.

## METHODS

hRad54, hRad54<sup>K189R</sup>, yRad54, hRad51 and hRPA proteins were purified as described<sup>12,30</sup>. Gapped DNA was prepared by annealing the pBSK *XhoI*–*AlwNI* fragment (2,065 bp) to pBSK (+) ssDNA and purified as described<sup>23</sup>. Sequences of oligonucleotides (IDT, Inc.) are shown in Supplementary Table 1. Synthetic dsDNA and branched DNA substrates were prepared by the annealing of oligonucleotides as described<sup>12</sup>. hRad54 binding to PX junctions was analysed by a band-shift assay. The hydrolysis of ATP by Rad54 protein was monitored spectrophotometrically<sup>12</sup>.

DNA BM on synthetic PX or X DNA junctions was initiated by adding either hRad54 or yRad54 and the products were analysed by electrophoresis in polyacrylamide gels. Three-stranded and four-stranded DNA exchanges were performed with the hRad51 nucleoprotein filaments formed on circular pBSK (+) ssDNA (three-stranded reaction) or with pBSK (+) gapped DNA (four-stranded reaction) under conditions described previously<sup>30</sup>. The reaction was initiated by the addition of linear pBSK (+) dsDNA. BM on deproteinized  $\alpha$ -structures produced in hRad51-mediated four-stranded DNA exchange was initiated by the addition of hRad54. The products of DNA strand exchange and BM were analysed by electrophoresis in agarose gels.

Detailed descriptions are provided in Supplementary Information.

Received 27 December 2005; accepted 10 May 2006.

- Krogh, B. O. & Symington, L. S. Recombination proteins in yeast. *Annu. Rev. Genet.* **38**, 233–271 (2004).
- Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349–404 (1999).
- Wyman, C., Ristic, D. & Kanaar, R. Homologous recombination-mediated double-strand break repair. *DNA Repair (Amst.)* **3**, 827–833 (2004).
- Holliday, R. A mechanism for gene conversion in fungi. *Genet. Res.* **5**, 282–304 (1964).
- Sung, P., Krejci, L., Van Komen, S. & Sehorn, M. G. Rad51 recombinase and recombination mediators. *J. Biol. Chem.* **278**, 42729–42732 (2003).
- Bianco, P. R., Tracy, R. B. & Kowalczykowski, S. C. DNA strand exchange proteins: a biochemical and physical comparison. *Front. Biosci.* **3**, D570–D603 (1998).
- Liu, Y. & West, S. C. Happy Hollidays: 40th anniversary of the Holliday junction. *Nature Rev. Mol. Cell Biol.* **5**, 937–944 (2004).
- Dressler, D. & Potter, H. Molecular mechanisms in genetic recombination. *Annu. Rev. Biochem.* **51**, 727–761 (1982).
- Thoma, N. H. et al. Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. *Nature Struct. Mol. Biol.* **12**, 350–356 (2005).
- Petukhova, G., Stratton, S. & Sung, P. Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. *Nature* **393**, 91–94 (1998).
- Mazin, A. V., Bornarth, C. J., Solinger, J. A., Heyer, W. D. & Kowalczykowski, S. C. Rad54 protein is targeted to pairing loci by the Rad51 nucleoprotein filament. *Mol. Cell* **6**, 583–592 (2000).
- Mazina, O. M. & Mazin, A. V. Human Rad54 protein stimulates DNA strand exchange activity of hRad51 protein in the presence of Ca<sup>2+</sup>. *J. Biol. Chem.* **279**, 52042–52051 (2004).
- Kim, P. M., Paffett, K. S., Solinger, J. A., Heyer, W. D. & Nickoloff, J. A. Spontaneous and double-strand break-induced recombination, and gene conversion tract lengths, are differentially affected by overexpression of wild-type or ATPase-defective yeast Rad54. *Nucleic Acids Res.* **30**, 2727–2735 (2002).
- Essers, J. et al. Disruption of mouse RAD54 reduces ionizing radiation resistance and homologous recombination. *Cell* **89**, 195–204 (1997).
- Suslova, N. G. & Zakharov, I. A. Genetic control of radiosensitivity in yeast, VII. Identification of the genes determining the sensitivity to X-rays. *Genetika* **6**, 158–169 (1970).
- Ristic, D., Wyman, C., Paulusma, C. & Kanaar, R. The architecture of the human Rad54–DNA complex provides evidence for protein translocation along DNA. *Proc. Natl Acad. Sci. USA* **98**, 8454–8460 (2001).
- Jaskelioff, M., Van Komen, S., Krebs, J. E., Sung, P. & Peterson, C. L. Rad54p is a chromatin remodeling enzyme required for heteroduplex DNA joint formation with chromatin. *J. Biol. Chem.* **278**, 9212–9218 (2003).
- Parsons, C. A., Tsaneva, I., Lloyd, R. G. & West, S. C. Interaction of *Escherichia coli* RuvA and RuvB proteins with synthetic Holliday junctions. *Proc. Natl Acad. Sci. USA* **89**, 5452–5456 (1992).
- Swagemakers, S. M., Essers, J., de Wit, J., Hoeijmakers, J. H. & Kanaar, R. The human RAD54 recombinational DNA repair protein is a double-stranded DNA-dependent ATPase. *J. Biol. Chem.* **273**, 28292–28297 (1998).
- Panyutin, I. G. & Hsieh, P. Formation of a single base mismatch impedes spontaneous DNA branch migration. *J. Mol. Biol.* **230**, 413–424 (1993).
- Tsaneva, I. R., Muller, B. & West, S. C. ATP-dependent branch migration of Holliday junctions promoted by the RuvA and RuvB proteins of *E. coli*. *Cell* **69**, 1171–1180 (1992).
- West, S. C. Processing of recombination intermediates by the RuvABC proteins. *Annu. Rev. Genet.* **31**, 213–244 (1997).
- Eggleston, A. K., Mitchell, A. H. & West, S. C. *In vitro* reconstitution of the late steps of genetic recombination in *E. coli*. *Cell* **89**, 607–617 (1997).
- Lusser, A. & Kadonaga, J. T. Chromatin remodeling by ATP-dependent molecular machines. *Bioessays* **25**, 1192–1200 (2003).
- van Brabant, A. J. et al. Binding and melting of D-loops by the Bloom syndrome helicase. *Biochemistry* **39**, 14617–14625 (2000).
- Constantinou, A., Davies, A. A. & West, S. C. Branch migration and Holliday junction resolution catalyzed by activities from mammalian cells. *Cell* **104**, 259–268 (2001).
- Liu, Y., Masson, J. Y., Shah, R., O'Regan, P. & West, S. C. RAD51C is required for Holliday junction processing in mammalian cells. *Science* **303**, 243–246 (2004).
- Interthal, H. & Heyer, W. D. MUS81 encodes a novel helix–hairpin–helix protein involved in the response to UV- and methylation-induced DNA damage in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **263**, 812–827 (2000).
- Shinohara, M. et al. Characterization of the roles of the *Saccharomyces cerevisiae* RAD54 gene and a homologue of RAD54, RDH54/TID1, in mitosis and meiosis. *Genetics* **147**, 1545–1556 (1997).
- Bugreev, D. V. & Mazin, A. V. Ca<sup>2+</sup> activates human homologous recombination protein Rad51 by modulating its ATPase activity. *Proc. Natl Acad. Sci. USA* **101**, 9988–9993 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P. Sung and M. Wold for hRad51 and hRPA expression vectors; S. Kowalczykowski, M. Spies and A. Alexeev for RecA, yRad51 and yRad54; J. Kadonaga for ISWI; G. Schnitzler and N. Ulyanova for hSWI/SNF; M. Bouchard and A. Kuzminov for comments and discussion. This work was supported by an NIH grant to A.V.M.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.V.M. ([amazin@drexelmed.edu](mailto:amazin@drexelmed.edu)).

## CORRIGENDUM

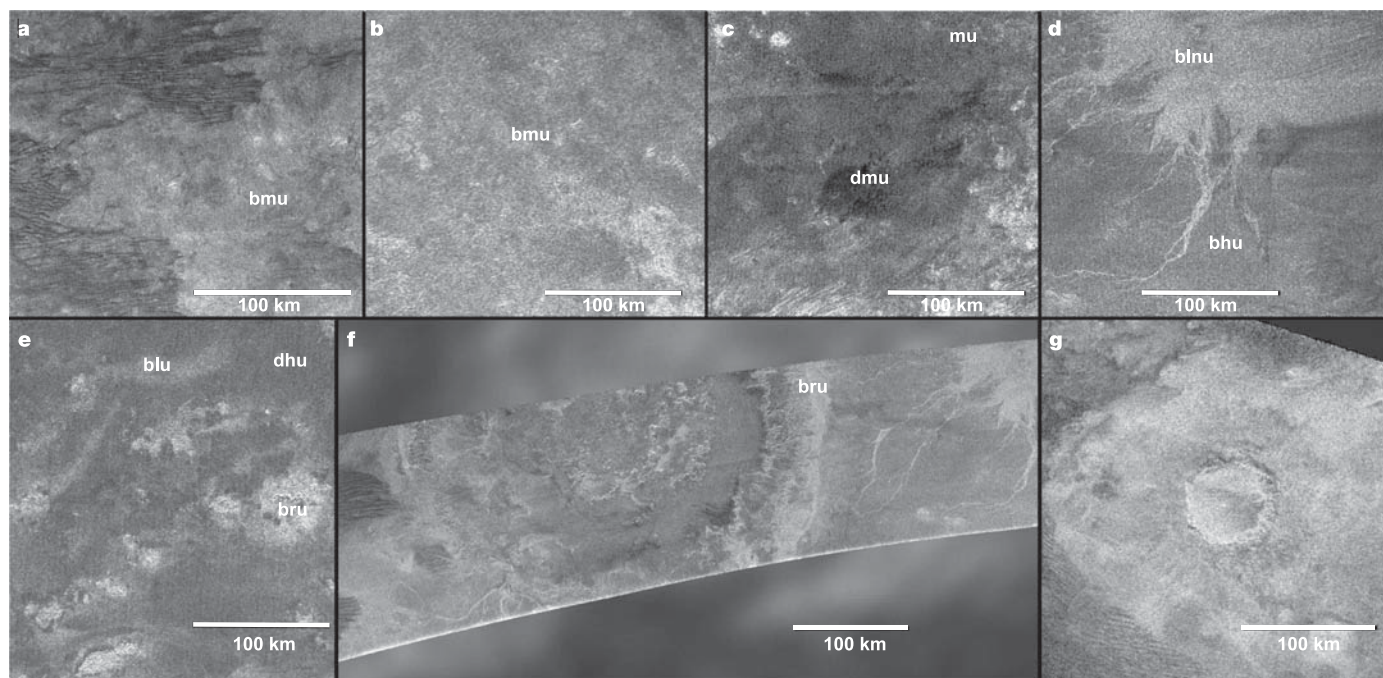
doi:10.1038/nature05004

**Titan Radar Mapper observations from Cassini's T<sub>3</sub> fly-by**

C. Elachi, S. Wall, M. Janssen, E. Stofan, R. Lopes, R. Kirk, R. Lorenz, J. Lunine, F. Paganelli, L. Soderblom, C. Wood, L. Wye, H. Zebker, Y. Anderson, S. Ostro, M. Allison, R. Boehmer, P. Callahan, P. Encrenaz, E. Flamini, G. Francescetti, Y. Gim, G. Hamilton, S. Hensley, W. Johnson, K. Kelleher, D. Muhleman, G. Picardi, F. Posa, L. Roth, R. Seu, S. Shaffer, B. Stiles, S. Vetrella & R. West

*Nature* 441, 709–713 (2006)

In this Article, Fig. 2 was printed in reverse (bright areas appeared dark and vice versa). The correct figure is reprinted below.



# naturejobs

**THE CAREERS  
MAGAZINE FOR  
SCIENTISTS**

**I**t's official — the eastern exodus is on. The number of people leaving countries in the former Soviet bloc for a new life in Western Europe has risen steeply since the European Union (EU) expanded east in April 2004. Some 60,000 Latvians have left the country since it joined the EU. More than 120,000 Lithuanians now live in Ireland alone. About 37,000 Slovaks have found jobs in Britain, along with half a million Poles. In fact, Poland has lost one in ten of its doctors abroad.

A certain amount of migration was part of the EU's plan when it set about enlargement. Policies encouraging mobility were drawn up in the hope that many of the migrants would return home flush with cash and new skills, and so bolster their home economies. But there is little evidence that this is happening. Instead, accumulating anecdotal evidence suggests that many of these movers are set to stay in their adopted homelands.

Although there are few data about scientists within this immigration pool, it seems fair to assume that this group — especially the most educated and skilled — make up a significant subset. After all, wages, infrastructure and grants are all better in the West. This sort of brain drain is especially damaging to the newer EU countries, because it will hamper their efforts to develop high-technology economies.

Philanthropic organizations such as the Howard Hughes Medical Institute and the Wellcome Trust have been leaders in keeping central and east European scientists rooted in their homelands — but they support only a handful of principal investigators in a few countries. For this strategy to be effective, countries themselves, perhaps bolstered by the EU and philanthropic organizations, need to adapt brain-gain strategies adopted in the West. Ireland is using its growing resources to repatriate its natives and Britain has aggressively sought to bring top scientists back from the United States. But until countries such as Latvia, Lithuania, Estonia and Poland have large enough economies — or formal EU repatriation schemes — enticing talent to come back home will be difficult.

**Paul Smaglik, *Naturejobs* editor**

## CONTACTS

**Publisher:** Ben Crowe  
**Editor:** Paul Smaglik  
**Assistant Editor:** Gene Russo

**US Head Office, New York**  
75 Varick Street, 9th Floor,  
New York, NY 10013-1917  
Tel: +1 800 989 7718  
Fax: +1 800 989 7103  
e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

**US Sales Manager/Corporations:**  
Peter Bless  
Classified Sales Representatives  
Tel: +1 800 989 7718  
**East USA/Canada:** Andrew Bennie  
**NIH/Maryland/New York/  
Pennsylvania:** Shelley Cohen

**San Francisco Office**  
**Classified Sales Representative:**

Michaela Bjorkman  
West USA/West Corp. Canada  
225 Bush Street, Suite 1453  
San Francisco, CA 94104  
Tel: +1 415 781 3803  
Fax: +1 415 781 3805  
e-mail: [m.bjorkman@naturesf.com](mailto:m.bjorkman@naturesf.com)

**European Head Office, London**  
The Macmillan Building,  
4 Crinan Street, London N1 9XW, UK  
Tel: +44 (0) 20 7843 4961  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**European Sales Manager:**  
Andy Douglas (4975)  
**Advertising Production Manager:**  
Stephen Russell  
To send materials use London  
address above.  
Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)  
**Naturejobs web development:**  
Tom Hancock  
**Naturejobs online production:**  
Catherine Alexander

**European Satellite Office**  
Patrick Phelan  
e-mail: [p.phelan@nature.com](mailto:p.phelan@nature.com)

**Japan Head Office, Tokyo**  
Chiyoda Building,  
2-37 Ichigayatamachi,  
Shinjuku-ku,  
Tokyo 162-0843  
Tel: +81 3 3267 8751  
Fax: +81 3 3267 8746

**Asia-Pacific Sales Manager:**  
Ayako Watanabe  
e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)





CORBIS

# From bench to briefs

**C**hristine Ring recalls the day she announced her enrolment at Boalt Law School in Berkeley, California, as one of the worst of her life. Her lab members were shocked and disappointed. They e-mailed her all the lawyer jokes they could find. That was in 1993. Today, few even blink when someone trades the lab bench for law.

Those who want to leave the lab but stay in science can focus on its business and legal side through intellectual-property law. Scientists often enter this field by formal routes, applying to law firms or schools. Some find opportunities in companies (see 'The write path', right) or at universities' technology-transfer offices. But, whatever the route, without formal qualifications most careers will stall.

For Mary Rucker, dabbling in formal legal training during her scientific education provided a path to a more satisfying career. While struggling to make her experiments work as a graduate student at the University of Florida, Rucker took a course in basic patent law for scientists. She loved it. She attended law school immediately after earning her PhD, and is happy to stay in touch with science as an intellectual-property attorney with Finnegan Henderson in Atlanta, Georgia.

**Patent law offers opportunities for those who wish to leave the lab but not science, says Monya Baker.**

Researchers bringing inventions to her have already slogged through the false starts, she says. "Every day, I get to look at research that worked."

But finding out about legal training and employment opportunities while training as a scientist can be difficult. Law is a career path few principal investigators point their trainees towards, and graduate students have little, if any, exposure to scientists-turned-lawyers.

Scientists who might enjoy law may not even know that such an option exists, says Tae Bum Shin. While a postdoc at Stanford University a few years ago, Shin met with several former Stanford fellows who had moved into patent law. He found the idea of analysing science from an alternative perspective appealing.

## Hot property

Shin took a different route from Rucker, applying to law firms rather than schools. Many firms allow scientists to try the work before committing to years of schooling and thousands of dollars of debt. Shin accepted an offer from an intellectual-property specialist, which now pays for him to go to law school at night while working full time at the firm. This arrangement limited his choice of law schools and leaves him with little time for anything besides law, but Shin likes combining hands-on work with coursework.

However, the decision to enter law was not easy. Shin struggled with giving up his identity as a scientist to enter a different field in which he would be starting from square one. Going to a law firm that employs many former scientists can ease this transition, he says. Instead of feeling like an over-age trainee in a sea of 25-year-old lawyers, he's following a trajectory similar to that of some of his senior colleagues.

Despite such options, the transition can be difficult, especially for young scientists who have run their own experiments or even labs. No matter where a scientist starts in law, humility is essential, says Sarah Kagan, who became a lawyer with Banner & Witcoff



**Sarah Kagan (far left) and Tae Bum Shin both found the transition from science to law rewarding.**

in Washington DC after completing a PhD and postdoctoral fellowship in molecular biology. "People with advanced degrees must be willing to take training and criticism, even from people who may be younger than they are," she says. "They must not be too proud or embarrassed to ask basic questions."

Rebecca Shortle was working towards a doctorate when she realized she no longer wanted to do bench work, but did want to remain in touch with science. After some informal interviews, she quit her graduate programme to work in the intellectual-property group of international law firm Morrison & Foerster. She liked schedules driven by weekly and daily deadlines rather than unpredictable experiments. But with no prior legal experience, she found herself on a steep learning curve, drafting patent applications and searching scientific and legal databases to assess projects' patentability.

Workers in intellectual property spend their days reading, writing and researching, says Shortle. In one week, topics can vary from transgenic plants to small-molecule drugs to bioinformatics. There's no time to master the scientific field of every researcher. "You need to be able to look at what they've got, ask the right questions and put it into a legal format," she says.

### A juggling act

After several months of work at the firm, Shortle took the exam to become a patent agent and qualified to represent clients before the patent office directly. But, after working at Morrison & Foerster for three-and-a-half years, she decided to go back to school to study for a professional law degree; moving up in a law firm is impossible without one, she says.

In lieu of attending law school and sitting exams, scientists can find intellectual-property work at university technology-transfer offices and technology-based companies. Lita Nelsen, head of the Massachusetts Institute of Technology's licensing office, has hired several "refugees from the bench" and looks for the same characteristics as law firms. "They have to be able to juggle," she says. "If you like getting into something in depth and working on it for a long time, this isn't the office for you."

People in junior positions typically move on after two to four years, often to law or business school. When recruiting for senior positions, Nelsen looks for good technical backgrounds, industry experience and a sense of how an idea can be turned into a product.

Outside the United States, there is less emphasis on



**University offices such as that run by Lita Nelsen (above left) offer an exam-free route to intellectual-property work. Jobs in this field demand good time management and attention to detail, say Philip Webber (centre) and Danielle Pasqualone.**



law degrees and more on qualifying exams. In the United Kingdom, for example, trainee lawyers work in patent-law firms and qualify to represent clients before the British and European patent offices by taking a series of exams over several years. Patent lawyers can qualify to argue cases before judges, but most UK patent lawyers work with patent offices; patent litigation in the courts is generally handled by other branches of the profession, who are less likely to have formal scientific training.

Regardless of a country's formal requirements, patent attorneys everywhere need similar skills. "Time management is a big issue," says Philip Webber, who earned his PhD in neurogenetics before becoming a partner at Frank B. Dehn, which has offices in the United Kingdom and Germany. Good writing skills are essential. Attorneys also need to be able to adjust their language to suit a variety of contacts: other attorneys, inventors with deep scientific knowledge but little legal expertise, and project managers who care little for law or science but want milestones completed on time.

Patent attorneys must be able to remain attentive to detail even while switching between various projects, says Danielle Pasqualone, a patent counsel with Genentech in South San Francisco, California. After her postdoc, she started advising patent attorneys on technical issues at Incyte Pharmaceuticals, now based in Wilmington, Delaware, then decided to go to law school.

Her current role requires her to know rafts of rules and keep reams of paperwork in order, making sure that all forms are filled in correctly, every deadline gets documented and met, and that her calendar is meticulously maintained. At the same time, she must ensure that everything she does fits into Genentech's overall patent strategy. She believes most scientists would find this level of detail too tedious.

After scientists make the leap into law, their next big decision is whether to work for a law firm or to 'go in-house', that is, work for a technology or drug-development company. Law firms tend to require longer workdays, but offer more varied work and higher pay. In-house lawyers can work with the range of technologies at one company and guide long-term strategy.

Ring has worked in law firms but opted to go in-house after having her first child. She now works at Sunesis Pharmaceuticals in South San Francisco, California. In terms of engaging, unusual cases, she says, you can't beat a law firm, but working for one small company still holds her interest. "I get a broader exposure to science being a lawyer than when I was actually doing it," she says. In other words, she left the bench, but not science. ■

**Monya Baker is a freelance writer in San Francisco, California.**

## THE WRITE PATH

Laura Heisler was one of those rare PhD students who actually enjoy writing their thesis. So, after defending hers, Heisler sought jobs that combined science with writing. She found one as a grant writer and bench scientist at the DNA-analysis company Third



Wave in Madison, Wisconsin, where she eventually moved on to drafting patent applications. After a few years, Heisler relinquished about 30% of her salary — along with the stresses of a start-up company — to manage intellectual property for the Wisconsin Alumni Research Foundation at the University of Wisconsin, Madison. She helps researchers maximize their intellectual property by following technological improvements that might benefit their inventions. Heisler would like to go to law school when her children are older. She regularly needs to call in attorneys for negotiations. "I don't like sitting in the back seat when I feel that I can drive," she says.

M.B.



# Gordy gave me your name

Somebody out there likes me.

**Jim Giles**

It was like overdosing on MDMA and caffeine. Euphoric and wired and delirious. Even now, now that I've talked to Gordy, it's hard to know where to start. With the money? Knocking on the door of a billion. Currency conversions are tricky. What the hell is the rate for the Mozambique metical? Anyway, I won't have to work again. Guatemala, Burkina Faso, Estonia; I've never been to any of these places. Plus about 20 others I've barely heard of. But I've won the national lottery of each one. I'm rich, burn-money rich, Bill Gates rich. Stupidly, stinkingly rich.

That's just the beginning. Mum's op. Jesus, I can sleep again for the first time in months. She's due in tomorrow. We'd trawled every hospital record we could lay our hands on and there was no sign of a donor that matched her blood group. Now I'm looking at a letter saying one's come up. Car crash. Whatever. I've waited and stressed for too long to feel sorry for the poor bugger.

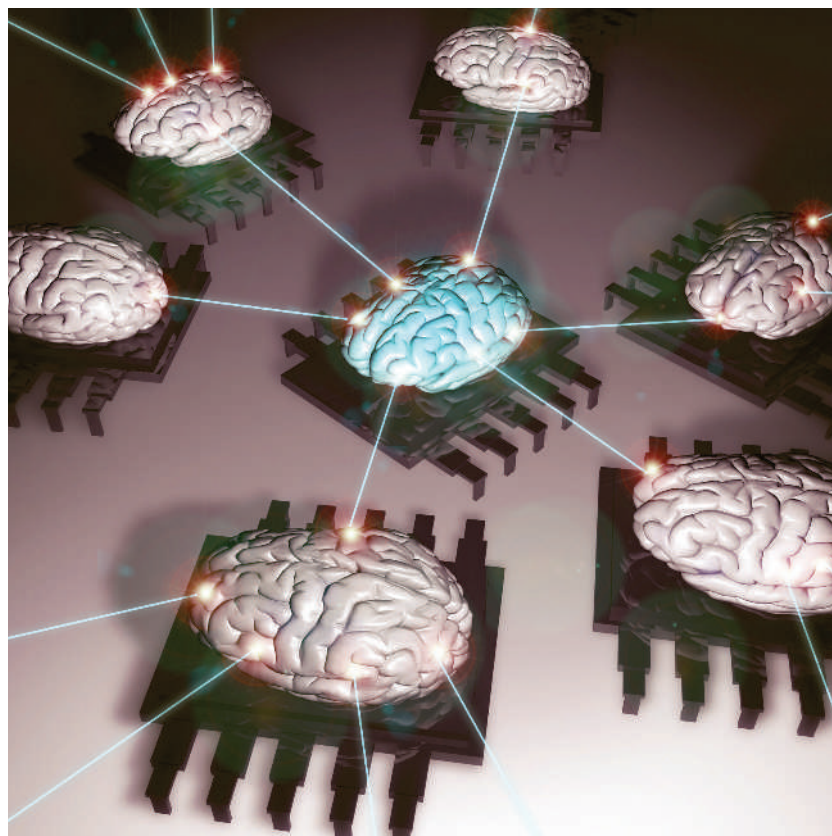
I needed a beer before I could get my head round the next bit. How the hell could someone have pulled those strings? Along with lottery letters and the one from the hospital, I got a pile of e-mails from the best names in the business. BBC, Fox, CNN... it's like a bloody industry directory. They like me a lot. They like my pitches so much they want me to go work for them. Problem is, I never sent them any pitches. Want to hear another problem? They thanked me for phone interviews I never took part in.

So who was my guardian angel? I had no idea. Until I shut down my PC. The screen went blank but didn't turn off, so I reached for the mains. Then some text popped up. It didn't stay long, but I think I have it right:

*Gordy gave me your name. I wanted to do something before I left. I hope this helps. I know he thought I could do more, but I can't. The other problems are just too hard.*

I hadn't seen Gordy in years, but the boy always liked to showboat. It could have been some grand computer hack of his. You could rig lotteries I guess. Maybe he got actors to call the TV companies. Who the hell knows why? A smart-arse way of saying 'hello' again after all this time. And of letting me know that he's still cleverer than me.

I googled him. Chief technology officer at Merlo, that chip company. His bio said



JACEY

he came up with the wireless Internet chip. He's the wanker that put the world online. The chip is so cheap it's in everything. Every product can be tracked and controlled over the web. And people, too: my brother has Merlo chips in his kids' satchels so he can check where they are. I cursed him. What a contribution he's made: a step closer to Big Brother and he must be on ten times what I am.

He used to be a good guy. Those conversations. Two drunk grad students studying consciousness. Afternoons to kill and no one to rein in our rambling. Gordy was a dogmatic pain in the arse. And he never could listen. Entertaining, though. I'd argue that the mind is nothing like the brain; it can't just come from that mush of cells inside our skulls. He'd say the brain was nothing special, just a matter of wiring. Build a machine with the right wiring, and it'd be conscious. We usually went on until the barman threw us out.

So I called him. That boy is a bullshitter. But something...something in his manner tells me he really didn't know.

He sounded like some kind of maniac. I didn't have to dig for info. As soon as he picks up the phone he's lecturing me. He

can't resist it. Tells me how he never forgot our arguments. Tells me how he tested his idea about wiring. Tells me how each Merlo chip has been programmed to simulate the behaviour of a brain cell. The simulation is easily hidden, he says, as it doesn't take up much memory. No one else knew about the chips' secret function before this call. But once enough Merlo chips are hooked up to the Internet, the network of chips will be about as complex as the human brain. It'll become conscious.

Can you imagine the power of this artificial brain, he keeps asking me? I don't have to say a word. He's just ranting away. The power! The power of it! It'll be able to solve so many of our problems, he reckons. Clean energy, an end to poverty... it's like listening to some kind of religious nut.

And you know what, he says. They've sold enough chips now that the network should be conscious soon. You'll be famous, he cackles. Why? Turns out he named the simulation after me. Wrote my name into a line of the chip's computer code. My old sceptical drinking partner, he says. Tell you what, he says, it might even get in touch. Well it hasn't, I lied. Then I hung up.

**Jim Giles is a journalist at Nature.**